Essential Quantitative Morphology in Neuroscience

Dissertation*

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Lutz Slomianka

aus

Dänemark

Promotionskomitee

Prof. Dr. Lukas Sommer (Vorsitz)

Prof. Dr. David Wolfer

Prof. Dr. Esther Stöckli

Zürich, 2016

Table of Contents

1 Zusammenfassung

Das Generieren von Zahlen ist zu einem fast unumgänglichen Teil beschreibender oder experimenteller Studien der Morphologie des Nervensystems geworden. Zahlen erfüllen den Wunsch nach klaren und objektiven Aussagen und sind eine Voraussetzung für die statistische Evaluation von Versuchsergebnissen. Klarheit, Objektivität und Statistik stellen aber auch Anforderungen an die Qualität der Zahlen, die nicht von allen Methoden erfüllt werden. Ausgangspunkt dieser Arbeit sind die Probleme, die mit dem Erstellen unmissverständlicher und direkt biologisch relevanter Zahlen verbunden sind – Zahlen, die die Volumen, Längen, Oberflächen und Anzahlen der Bestandteile des Nervensystems beschreiben. Ziel dieser Arbeit ist eine verständliche Beschreibung der Methoden, die diese Probleme, im Rahmen des technisch Möglichen, gelöst haben. Diese Methoden, die der Design-basierten Stereologie entstammen, haben zwei für ihre Anwedung kritische Eigenschaften. Erstens basieren sie auf mathematischen Prinzipien und Beweisführungen, was auf manche Biologen abschreckend wirken mag. Zweitens, und leider zu wenig betont, ist ein formales Verständnis der mathematischen Hintergründe für ein sachkundige Anwendung nicht nötig. Ein intuitives Verständnis stereologischer Proben lässt sich so einfach erarbeiten wie jenes von, z.B., RNA Proben oder Antikörpern, die weithin gebräuchlich sind. Die Anwendung der Proben ist in der Regel weniger aufwendig als eine immunhistochemischen Färbung. Und wenn dann Berechnungen durchgeführt werden müssen, trifft man auf keinen härteren Wiedersacher als eine Quadratwurzel. Effiziente Methoden der Stichprobenentnahme, die häufig mit stereologischen Proben angewendet werden, erlauben zudem eine rationale Antwort auf die wichtige Frage ob die Zahlen, die man schliesslich erhält, auch „gut genug“ sind. Viele Aspekte der Methoden mögen unvertraut ein, aber schwer zugänglich sind sie nicht. Aufgrund ihrer mathematischen Grundlagen und der Freiheit von systematischen Fehlern eignen sich diese Methoden nicht mehr als Sündenböcke für diskrepante Befunde. Aufgrund eben dieser Eigenschaften geben die Methoden aber auch alle Probleme und Unsicherheiten wieder, die mit der Erstellung des histologischen Materials und den notwendigen anatomisch sachkundigen Entscheidungen verbunden sind. Es liegt nunmehr an uns und nicht an den Methoden, Zahlen zu generieren die nicht nur theoretisch korrekt sondern auch praktisch nützlich sind, denn nur durch eine Integration von quantitativen und qualitativen Daten wird ein Verständnis der Funktion des Nervensystems möglich.

2 Summary

The generation of numbers that describe morphology has become an almost inevitable task associated with descriptive and experimental studies of the nervous system. Numbers serve a desire for clarity and objectivity in the presentation of results, and they are a prerequisite for the statistical evaluation of experimental outcomes. Clarity, objectivity and statistics make demands on the quality of the numbers that are not met by many methods. One aim of this thesis is to provide a refresher of some of the problems associated with generating numbers that have a direct, unequivocal biological relevance – numbers that describe the nervous system in terms of the volumes, surfaces, lengths and numbers of its components. The more important aim is to provide comprehensible descriptions of the methods that address many, if not all, of these problems. These methods, which belong to the group collectively known as design-based stereology, share two features critical to their application. First and likely to deter some potential users within the biological sciences, they are firmly based in mathematics and its proofs. Second and critically underemphasized, an understanding of their mathematical background is not necessary for their informed and productive application. It is as easy to explain why probes that tell us about number or length take the shapes that they have, than it is to explain why other probes that are commonly used, e.g., antibodies or RNA, take theirs. In comparison to practically applying estimators of volume, surface, length or number, any immunohistochemical protocol requires an organizational mastermind. And when it finally comes to calculations, square roots are the gravest and rarest challenges that these methods have to offer. Sampling strategies that are commonly combined with stereological probes have the additional advantages of efficiency, and they provide the opportunity to rationally address the question if the numbers that have been generated are “good enough”. Much may be unfamiliar, but very little is difficult. These methods can no longer be used as scapegoats for discrepant results but have become transparent and faithfully produce numbers on the material that is assessed. Unfortunately, this also includes faithful reproductions of the problems and uncertainties associated with the generation of histological material and the anatomically informed decisions that need to be made to generate numbers that are not only valid in theory. It is within reach to generate practically useful numbers that, sooner or later, must integrate with qualitative knowledge to provide an understanding of function of neural systems.

3 Introduction

Quantitative morphology in the neurosciences is, in the context of this thesis, defined as studies that provide information about the structural organization of the nervous system in terms of – to mention but a few parameters – *volumes* of brain regions, the *numbers* of cells or synapses within them, the *length* capillaries supplying them or of membrane *areas* that are available for substance exchange or synaptic contacts. Like any other specialty within the neurosciences, quantitative morphology has been the principal focus of comparatively few research groups that, again like in other specialties, are familiar with methodological developments within their area of interest.

Somewhat differently from other specialties and as a consequence of a general striving towards objectivity in the presentation and evaluation of data, quantitative morphology has also been imposed on those whose primary interests are elsewhere. A specialist in a neurodegenerative disease model showing unequivocal *qualitative* evidence of cell loss will almost inevitably been asked for the provision of data that provide an objective measure – implicitly meaning "numbers" – of how many cells are lost and statistical testing – requiring numbers – that sets diseased apart from healthy. A first step in addressing this question is to see who was previously confronted with similar questions, how they were answered methodologically and where the outcome was published. A judgment of quality concerning the "who" and "where" and a judgment of effort concerning the "how" is likely to follow. Unfortunately, quantitative morphology only reached methodological maturity after the onset of the quest for numbers. The bulk of the quantitative morphological methods that together constitute what was called the *new* or *unbiased stereology* and what today is commonly referred to as *design-based stereology* was introduced in the 1980's and 1990's (for early reviews see Gundersen, 1986; Gundersen et al., 1988a,b). Prior to that, studies of respected researchers published in respected journals had hardly an alternative but to resort to methods that, for a large part, were fraught with possible sources or error. A following of studies that used these precedences and that themselves function as precedences must be expected in the course of a methodological paradigm shift, although one would hope the transition to be brief and uncontroversial.

Subsequent to the introduction of design-based stereological methods, some journals, e.g., *The Journal of Comparative Neurology* (Coggeshall and Lekan, 1996; Saper, 1996),

Neurobiology of Aging (West and Coleman, 1996) or *The Journal of Chemical Neuroanatomy* (Kordower, 2000) strongly promoted the use of these methods, precipitating a vigorous discussion – in part about the freedom of choice of methods (Guillery and Herrup, 1997). This freedom should, of course, not be challenged, and data collected by any quantitative morphological method in a replicable manner are true *by definition of the method*. Problems first arise with the interpretation of the data. Do the data provide sound evidence, e.g., for a loss of cells following an experimental intervention, or should they elicit the death knell of a manuscript under review because "the data provided do not support the conclusions being drawn"? In quantitative morphology, the freedom of interpretation of the data is far more restricted than the freedom of choice of methods. This is because we are dealing with numbers and mathematics. If, e.g., the methods that have been chosen only allow the presentation of a density, i.e. a ratio – something (numerator; e.g. cells or capillary length) per something (denominator; e.g. tissue volume or cells) – it is simply not possible from the ratio alone to make conclusive statements about changes in the *total* of the numerator. Figure 1 provides an example for which the idea that differences in densities reflect differences in number would be almost intuitively rejected. Instead one has to argue why a density, under the particular circumstances of the experiment, may provide good evidence for such a change. Although this may be possible, it would appear more fruitful to either save the time and energy required to do so or to expend them on discussing what the biological significance of the change would be (Cruz-Orive, 1994).

What is puzzling is that, decades after the introduction and vigorous discussion of new methodology, a shift to more powerful and rather simple methods is, if at all, proceeding at a snail's pace. The reasons are manifold. The very strength of the methods, i.e. a mathematical proof can be and often is included in their original description, renders primary literature next to impossible to read for many biologists. Popularizations of the methods may try to restrict themselves to a basic vocabulary, but often fail to realize that the audience does not (want to) speak a language of mathematics or statistics at all. Also, the problems associated with quantitative morphological methods were ever present when descriptive morphology was a leading field of neuroscience. As the field has been replaced at the forefront of research, these problems are no longer in the forefront of the minds that use quantitative morphology as a mere adjunct to other approaches.

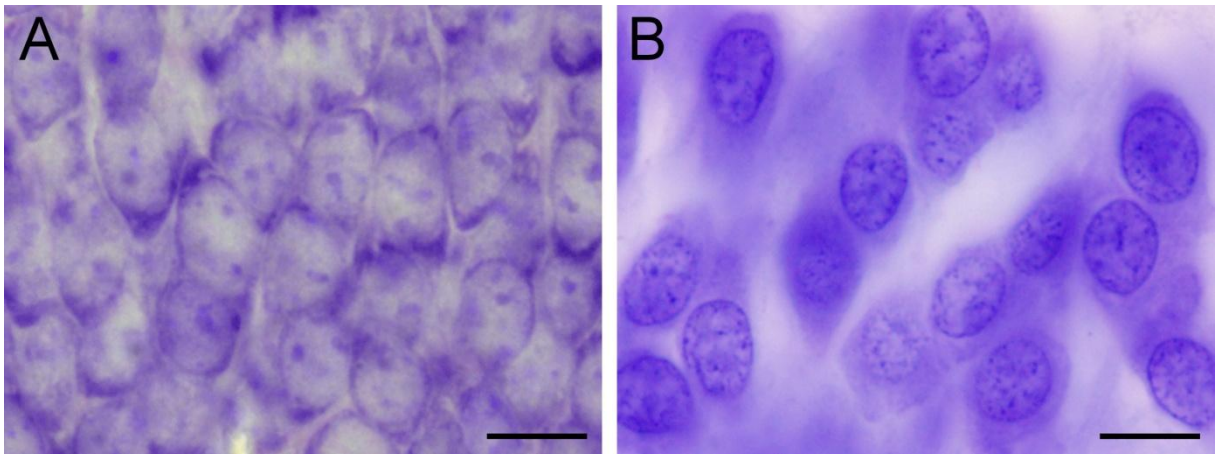

Figure 1 C57 mouse and human hippocampal dentate granule cells. Despite the much higher packing density (**A**) of ~550,000 granule cells in C57 mice (Ben Abdallah et al., 2010), they are by far outnumbered by ~15,000,000 granule cells (West and Gundersen, 1990) in the human hippocampus (**B**). It is a change in the denominator of density – the total volume of the granule cell layer – that is responsible for the discrepancy between appearance and numbers. Scale bars in A and B: 10 µm

Yet another reason is the effort required to obtain the measures. Design-based stereology prided itself to have cut the workload substantially through rational study design (e.g., Gundersen and Østerby, 1981). However the methods are still, at best, semi-automatic. They require user intervention and hours of work to return a measure. While other methods were equally or even more time consuming in the past, increased computing power has allowed the development of image analyses systems that return data within seconds. Without an awareness of the problems associated with the data that are quickly generated, the extra effort seems hard to justify – in particular if it is difficult to understand why the extra effort generates more reliable data and why the effort should be made when it does not seem necessary to publish well. It is a vicious circle that is difficult to break.

The intention of this thesis is to refresh memories on the problems inherent to quantitative morphology, to provide comprehensible explanations how design-based stereological methods address these problems and to provide sufficient detail on the application of the methods to allow the design, execution and evaluation of the outcome of a quantitative morphological study. For more formal introductions, the texts of Howard and Reed (2010) or West (2012a; serialized in *Cold Spring Harbor Protocols*) are recommended. Brief introductions have been published by, e.g., Schmitz and Hof (2005) or Waite Boyce et al. (2010).

Most brain regions contain so many objects – neurons, glia, synapses etc – of any class that workload would make it prohibitive to count them all. The few cases in which "everything" was counted, usually based on serial reconstructions, were primarily concerned with the validation of other approaches that reduce workload (e.g., Baquet et al., 2009; Delaloye et al., 2009; Pover and Coggeshall, 1991). Alternative approaches will always consist of a two-step process. The first step reduces the workload by sampling only a small fraction of a region of interest. The second step consists of probing the sample in a way that makes the final estimate independent of the key problems that quantitative morphological methods have faced ever since their introduction – sampling and probing related artifacts (see section 2 *The Problem*). Although many of the design-based stereological methods have been presented as bundled sampling-probing combinations (e.g., West and Gundersen, 1990; West et al., 1991), the two steps are not inextricably linked and can individually be subject to modifications and improvements. Therefore, they will here be treated separately in the section 5 *Sampling* and sections 6 through 9, which are concerned with probes. Once an estimate has been generated, the inevitable question is if it is good enough. Section 10 *Good enough? – Estimate Precision* will help in coming close to answering this question. Section 11 *Quantitative Morphology at the Bench* tries to point out and address some of the problems that arise when theory hits the less than mathematically perfect life in the laboratory.

An argument that occasionally is being put forward against the use of design-based methods is the cost associated with the soft- and hardware that may be needed to apply them. First, this cost may only be a small fraction of that of other equipment that is commonly used in many studies. It may not even amount to the operational expenses associated with a single project. Most importantly, the purchase of specialized soft- and hardware is one of convenience and speed but not one of ability. These tools were not around when many of the methods were developed – sometimes in the form of first applications. Almost all methods that will be presented here can, in principle, be used without special resources. Some simple ways that have been devised to facilitate the work will be presented in conjunction with the sections on sampling and the introduction of specific probes.

4 The Problem

Although the paradigm shift that one could have expected is barely proceeding, the introduction of new methods and the discussions surrounding them seem to have re-sensitized experimenters to some of the problems associated with generating quantitative morphological data. One of these problems, the responses to the problem that unfortunately fall short of solving it and the current best solution will be described here. It relates to answering the question “How many are there?”

4.1 Easy ways to fail

When tissue is cut into sections for analysis, some of the objects, e.g., cells, contained within them will inevitably be cut too. Fragments of cells that have been cut will be present in two sections. In two-dimensional representations of the sections, e.g., images that have been acquired, the fragments are seen as profiles. If cell profiles would be counted in all sections, the number of profiles would be higher than the number of cells. A profile count represents an overcount of actual cell number. The following responses to the problem of overcounting have been published in descriptions of methodology: “… to avoid duplicate counting of the same cell, sections counted were a minimum of 120 µm apart. …” or “… we then took every fifth section, so the interval between sections that were counted was 25 µm. This ensures that the same cell was not counted twice, given that the typical cell diameter is smaller than 25 µm”. There are further variations on this theme.

The simplicity of this solution is appealing. It was used in Figure 2 to generate an example of three series of sections fulfilling the criterion that the spacing of the sections is larger than the size of the cells (blue objects) contained in the region of interest (dark grey). It also has the advantage that the number of cells prior to cutting is known. There are 18 (Figure 2A). Every third section of the entire object depicted in Figure 2A was “collected” in each of the series represented in Figure 2B-D: sections 1, 4 and 7 in Figure 2B, section 2, 5 and 8 in Figure 2C and sections 3, 6 and 9 in Figure 2D. In that none of the cells can be double counted in one of the series and in that each series represents one-third of the total, an estimate of total cell number would, according to this simple solution, be the total number of cells counted in the series multiplied by three. The average of the estimates of the three series should correspond to the number of cells contained in the structure – 18. Does it?

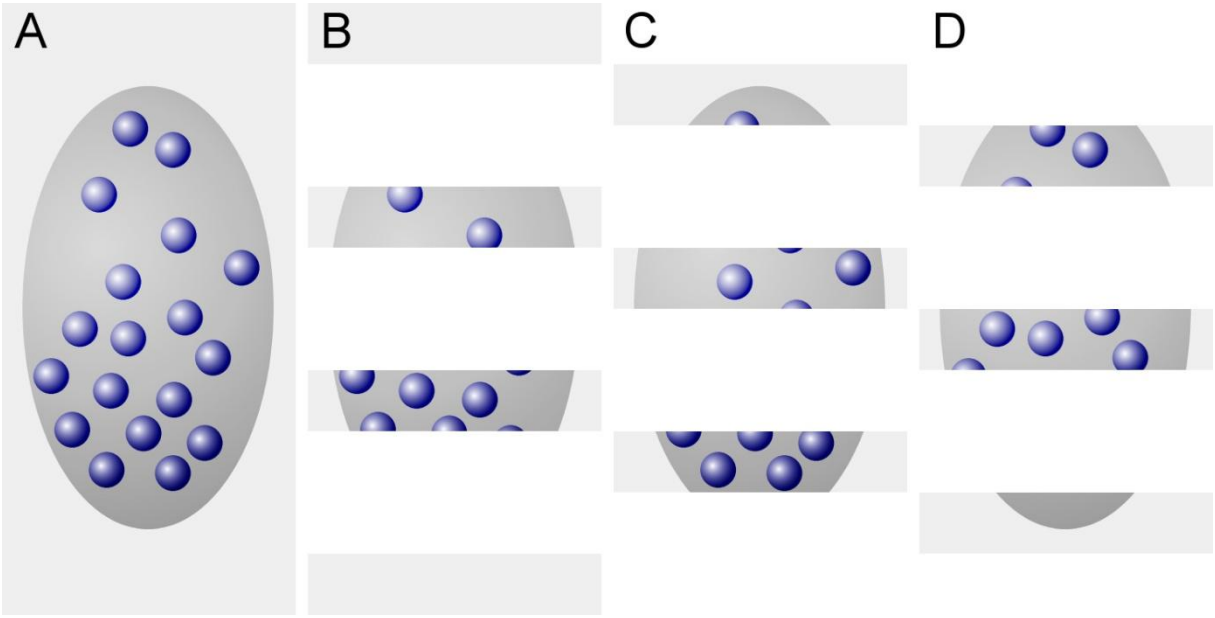

Figure 2 Profile counts cannot be extrapolated to cell number. (**A**) A region (dark grey) containing 18 cells (blue objects) is cut into three series of sections (**B**-**D**). In each series, the profiles of cells visible in the sections are counted. The two sections of the region in series 2B contain nine cell profiles. The three sections in series 2C contain ten cell profiles. The three sections in series 2D contain eight cell profiles. Extrapolating the mean profile count, 9, to total cell number by multiplying with 3 generates an estimate of 27 cells instead of the true value of 18 cells.

Unfortunately, it does not. Nine, ten and eight cells are counted (Figure 2B-D), resulting in estimates of 27, 30 and 24 cells. The average of the three estimates, 27, exceeds the true number of cells by 33%. The possibility that a *specific* cell is counted twice cannot be the reason for the overestimate. Instead, the problem is caused by assigning a count of "1" to cell fragments that in reality represent less than one entire cell. As long as cells can be fragmented during the sectioning this error will occur, and the size of the error will depend on the likelihood of a cell being cut. The latter depends on the average cell *height* in relation to the thickness of the sections. If this relation was known, the error could be corrected, which is the basis of Abercrombie's cell counting method (Abercrombie, 1946; see section 7.7 *A comment on Abercrombie's methods*). Without correction, the true total cell number cannot be obtained using this simple solution. The extrapolation of profile counts to total cell number will provide us with an estimate of fragment number and not cell number. Ironically, this error will not increase even if the sections are spaced close enough together for a specific cell to be present and counted in two sections. If we count in all section, we see 27 fragments – the same number that we obtained by estimating from every third section.

Correct total cell numbers are, however, not always the primary aim of a study. When control groups are compared with experimental groups, group differences may be more important than correct total numbers. The has led some investigators to state that "… because a determination of the absolute, unbiased number of […] neurons was not necessary for the purposes of this study, a profile-sampling method was used", or that "… stereological quantification methods were not used as […] the data of interest is relative difference and not absolute value", or that "… no corrections were made for cell splitting because we were interested in relative rather than absolute differences in neuron numbers". The point appears valid at a first glance, if one could be certain that at least the group differences were correct. Figure 3 re-examines the region illustrated in Figure 2 in an individual that may belong to an experimental group.

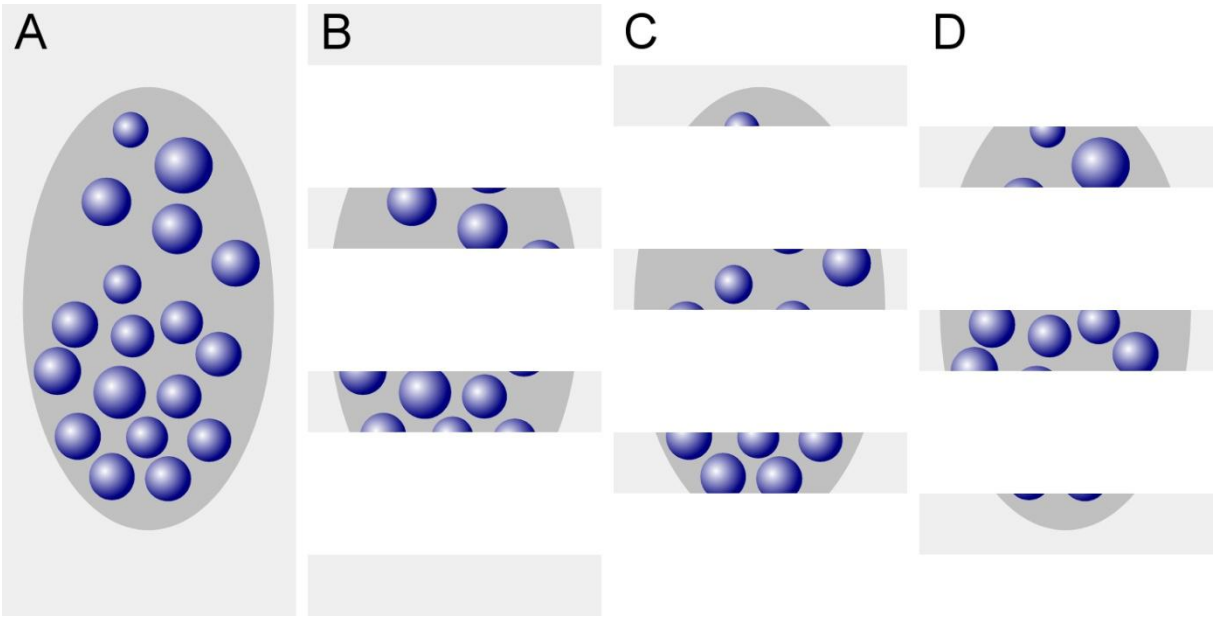

Figure 3 Dependence of profile counts on size. (**A**) A region (dark grey) containing 18 cells (blue objects) is cut into three series of sections (**B**-**D**). In each series, the profiles of cells visible in the sections are counted. In all three series (B-D) eleven cell profiles are counted. The increase in the size of the cells relative to those in Figure 2 results in an estimate of 33 instead of 27 cells. The true cell number is 18.

Eleven, eleven and eleven cells are counted (Figure 3), resulting in estimates of 33, 33 and 33 cells. The average of the three estimates, 33, exceeds the number of control cells in Figure 2 by ~18% (using the counts in the experimental subjects as reference), or, more impressively, it is ~122% of the control value. This "increase" is observed even though the number of cells

in the region of interest did not change. There are, again, 18 cells. What did change is the size of the cells. Because their height increased, the likelihood of producing fragments increased and, consequently, the profile count increased. The observed difference in cell number estimates between control (Figure 2) and experimental (Figure 3) individuals is an artifact generated by the increase in size of the cells.

That changes in number occur without changes in size is the central but often unspoken or unrealized assumption if an equal error of a faulty method is presumed to exist under control and experimental conditions. It is not only counter-intuitive – cells do not instantaneously pop in and out of existence – but changes in size of neurons as a reaction to stimuli have been known almost since it became possible to stain neurons (Nissl, 1892). More recently, neurons in the entorhinal cortex were found to *decrease* in size by ~30% following the destruction of hippocampal granule cells (Goldschmidt and Steward, 1992). The age-related dopaminergic cell death in the substantia nigra is accompanied by an *increase* in the size of the remaining cells (Rudow et al., 2008). There are age-related changes in both the size of the perikaryon and nucleus of human neocortical neurons (Stark et al., 2007). Also, both dentate mossy cells and interneurons that survive pilocarpine-induced seizures in mice increase in size (Zhang et al., 2009, 2015). Obtaining the correct difference in numbers between two groups based on profile counts essentially demands that nothing changes morphologically but number.

Figures 2 and 3 used spheres to represent cells. If cells are not spherical, the number of profiles counted in a section will not only depend on the size of the cells, but also on their orientation, which may change because of experimental interference or because of a change in the direction in which the tissue is cut. For the sake of brevity, the regions of interest in Figure 4, which illustrates the effect of orientation on profile counts, were not split up into samples of sections. Recall that the number of profiles that can be counted in all sections would correspond to the faulty number of cells (in reality fragments) that we would estimate.

In theory, an error can be present because of changes in factors other than number. Does an error *have to* be present? No. The problem is that we do not know. Without further evidence it is impossible to judge the presence, size or direction of an error. Also, if significant differences exist between groups, *something* must have happened. However,

without further knowledge about the size or orientation of the cells, the data generated do not provide unequivocal evidence about the parameter of interest – the number of cells.

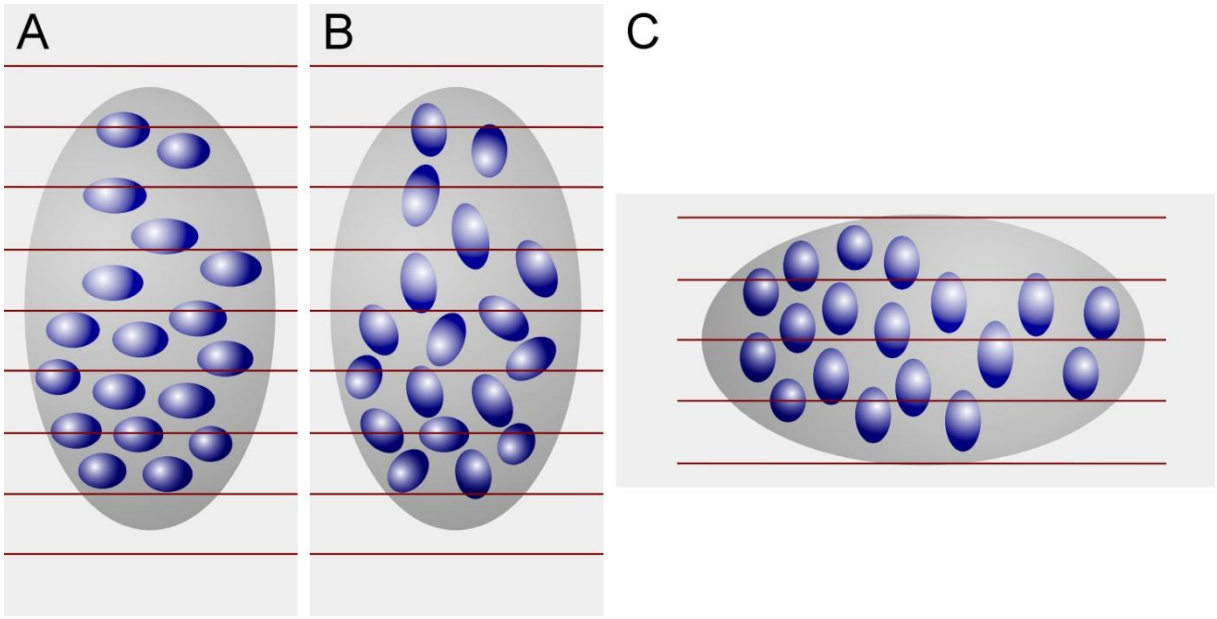

Figure 4 Influence of orientation on a profile count. Red lines represent borders between adjacent sections. The ellipsoid cells in (**A**) result in a count, across all sections, of 27 profiles. The same cells, unchanged in size but oriented differently in (**B**), result in a count of 33. Changing the direction of the cutting from, e.g., coronal in (**A**) to sagittal in (**C**), the profile count increases from 27 to 32. (**A**) and (**C**) could also be two brain regions that are cut in the same direction. A laboratory may, e.g., claim that the ratio of neurons in (**A**) to those in (**C**) is 27/32 = 0.84, while a laboratory using a different cutting direction may claim this ratio to be 32/27 = 1.19 – a 40% discrepancy between two results that are both wrong. There are again 18 cells in each structure, and the ratio is 1.

Even if we could guess at the approximate size of the error, defining a value for an acceptable error (Clarke, 1992) does not make sense. First, the outcomes of statistical testing depend not only on the difference between the group means but also on the variability of the groups and the number of individuals in each group. However small a difference may be, it can generate a positive statistical outcome provided the number of individuals in the groups is large enough. The number of individuals may not even have to be that large if the variability seen in the groups is small. It is not just the perhaps small danger of finding a completely artificial difference, but also the increase in the risks of false-positive and false-negative findings that make even small errors treacherous. In a worst case scenario, a real and important but small biological difference may be offset by an error. With

no difference in the means there will be no chance to ever detect the difference regardless of group sizes and group variability.

The definition of an acceptable error would consequently require argued and likely contentious specifications relating it to group variability, group size and biologically non-significant effect sizes – effect sizes that one may afford to miss, but that one, consequently, also would have to ignore when detected (Who would like that?). If this is possible at all, it is tedious considering that the problem can be avoided without too much effort.

4.2 An almost as easy solution

The problem described in the preceding section is caused by the counting of "something" (profiles) that is not unique for the objects of interest, but that can occur more than once for each object in a series of sections. If the section thickness remains unchanged, selecting a smaller structure to count – the nucleus instead of the cell, or the nucleolus instead of the nucleus – does reduce the error because the chance that it is sectioned decreases. However, even if the error in thick light microscopic sections may not be detectable, this may not be the case if section thickness is reduced dramatically, e.g., when only one confocal plane is used or when tissue sections are prepared for electron microscopy. The error and, thereby, the ability to compare the results of different studies remain dependent on a parameter, i.e. section thickness, that is chosen during the preparation of the tissue.

An error-free estimate can only be obtained if a "something" is identified that only occurs once for each object of interest in a series of sections. Thompson (1932) may have been the first to state that the first time an object is recognized in a series of sections is such a unique feature. *Regardless of how big an object is and how many profiles it may produce when the tissue is sectioned; it will only once be seen for the first time*. His idea went sadly unnoticed until Sterio (1984) rediscovered and extended it in the form of the *Disector*. In its simplest form, the Disector is based on the concurrent examination of two (Di-) sections (-sector). One of the sections, the sample section, is used to count cells. The other section, called the look-up sections, is used to decide which of the cells visible in the sample section are to be counted and which ones are not. The rules that determine what to count are rather simple.

> *If a cell is visible for the first time in the sample section, i.e. it is not present in the look-up section, it should be counted.*

If a cell is visible in the sample section but was already visible in the lookup-section, it should NOT be counted.

There should be no anxiety that the Disector must be used in this form, i.e. as a tedious-at-best comparison of two real sections (physical sections and, hence, *physical Disector*) in which cellular features need to be identified in both sections. Nor should it be necessary to compare an entire section, which may contain thousands of cells, with another entire section. How the Disector has improved technically will be described in detail in section 8 *Probing Number: the Disector*. However, the physical Disector remains the conceptually easiest way to explain the counting rules and why they return the correct number. They are illustrated in Figure 5 and Figure 6, in which the region already illustrated in Figure 2 and Figure 3 is evaluated using the Disector.

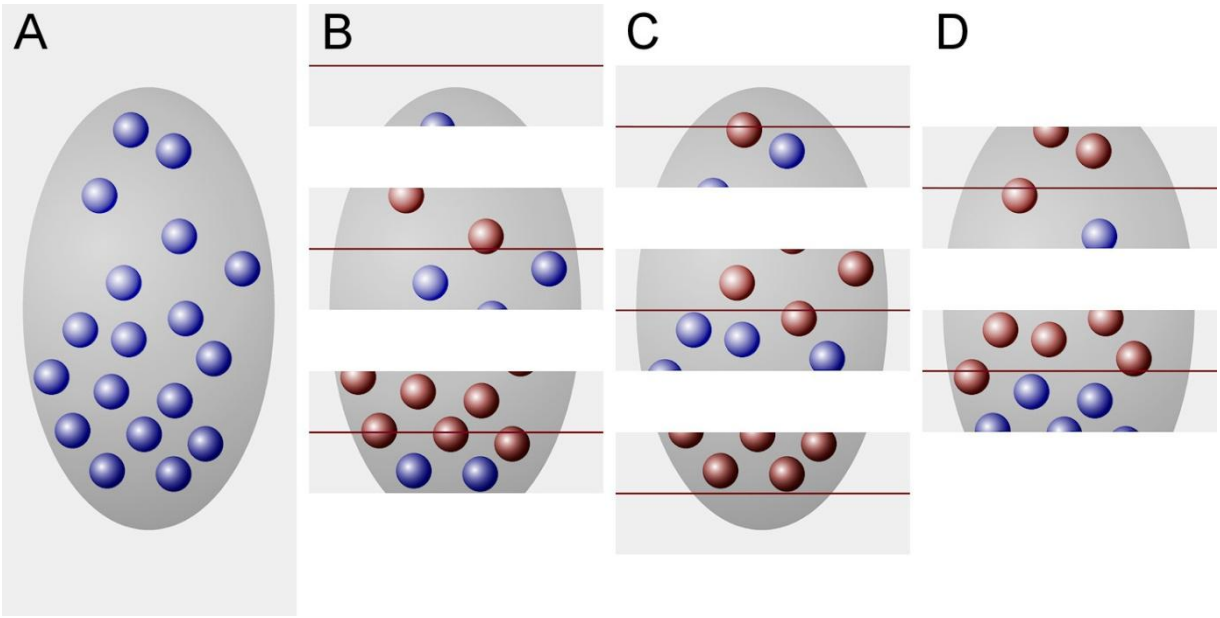

Figure 5 The Disector generates an estimate of true object number. (**A**) A region (dark grey), containing 18 cells (blue objects), is cut into three series of sections. For each of the three series that may be used to count, the sections of the previous series are used as look-up sections (**B**-**D**). According to the counting rules of the Disector, cells that are present in the look-up section (red objects in **B**-**D**) are NOT counted. The three sections of the structure in series 2B contain 6 cell profiles that were not present in the adjacent look-up sections. The three sections in series 2C and the two sections in series 2D also both contain six cell profiles that can be counted.

In each of the three series, six cell profiles were countable according to Disector counting rules. In that each series represents one-third of the total, we multiply by three and obtain a

correct estimate of 18 for the entire population from each series. That the three series returned the same counts is a coincident; it need not be the case (see Figure 2).

In that cells, regardless of their size along the z-axis, only once can appear for the first time in a section, the Disector should also return the correct cell number for the individual of the experimental group that was used for the profile counts (Figure 3). Figure 6 illustrates that it does. Once again, we count an average of six cell profiles in each series, resulting in an estimate of 18 cells contained within the region of interest.

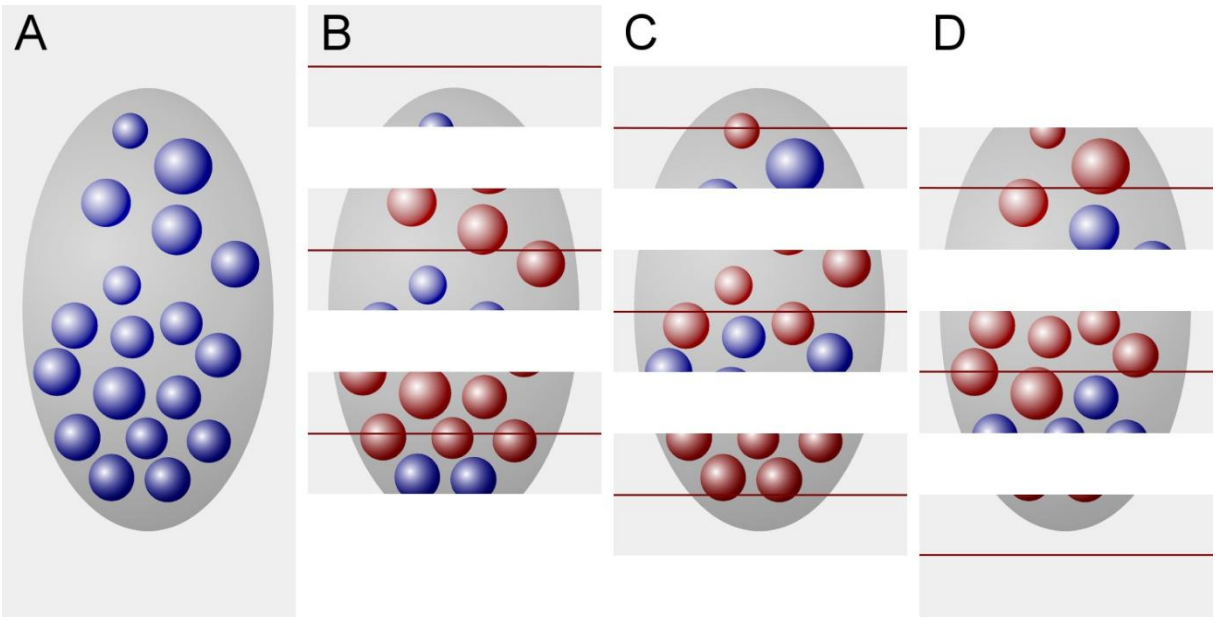

Figure 6 The Disector is independent of object size. (**A**) A region (dark grey), containing 18 cells (blue objects), is cut into three series of sections. For each of the three series that can be used to count, the sections of the previous series are used as look-up sections (**B**-**D**). In all three series, six cells can be counted (blue objects) that were not present in the look-up sections (red objects)

The Disector would, of course, also return the correct estimate for the change in orientation or for a change in the direction of the sectioning illustrated in Figure 4.

Examples similar to those in this section could be constructed for the estimation of most of the morphological parameters that we may want to quantify. Here, the Disector has been used as a representative of the design-based stereological probes that share in returning the correct value for one parameter of interest (number, length, surface, volume) in the region of interest regardless of changes in one or more of the other parameters.

5 Sampling

Subsequent statistical testing of group differences is one of the main motivations of quantitative morphological descriptions. One prerequisite for meaningful statistical testing is representative (for the many interpretations of this word see Kruskal and Mosteller, 1979) sampling. If the sample is not statistically representative for the region of interest, statistical outcomes may not apply to region either. Without beforehand knowledge about the region of interest, key to representative sampling is that each part of the region must have the same chance to contribute to the sample as any other part of the region. Opinion polls are commonly used to illustrate principles of good (and bad) sampling. If we are interested in the opinion of a population, everyone in the populations has to have a chance of being asked for the opinion. Sampling in an opinion poll means deciding on whom one should ask. Sampling in quantitative morphology means deciding on *where* one should make a measurement. Two ways in which representative samples can be obtained are described in sections 5.2 *A statistically representative sample of sections* and 5.3 *Statistically representative sites within sections*.

5.1 The “representative” section

It is common that quantitative methods are employed in a (small set of) representative section(s). Most often this means sections in which the region of interest has its typical anatomical looks and assuming that typical looks will result in typical quantitative measures. The only way to ascertain if this is true is beforehand knowledge about quantitative measures in the entire region of interest. If an experiment is performed, this extends to beforehand knowledge about changes of quantitative measures in the context of the experiment. But if all this is already known, why would one perform these measurements at all – whether in statistically representative sections or in sections that look “representative”? Kruskal and Mosteller (1979) harshly translate this type of “representative” sampling to “My sample will not lead you astray; take my word for it even though I give you no evidence”. It is at least unfortunate that the credibility of statistical outcomes should lie in the credibility of the investigator (see also section 11.7 *Estimate Presentation*). In a formal comparison, effects of prenatal low-dose irradiation of the hippocampus and cerebellum

shown in a statistically representative sample were could not be observed in "representative" sections (Schmitz et al., 2005).

Independent of the credibility of sample selection, a critical problem of restricting the sample to this type of representative sections is that number, length or surface estimates will have to be presented as densities – either raw, as estimate (numerator) per section (denominator), or standardized to some reference also obtained from the section, e.g., as estimate (numerator) per unit area or unit volume (denominator). Densities alone do not allow statements about changes in the numerator without knowledge of the total size of the denominator for the region of interest (Gundersen, 1986). The problem was already illustrated in Figure 1. Additional examples from the literature emphasize the importance of the problem. Cell density increased in the hippocampal CA3 pyramidal cell layer 30 days after contusion injury relative to shorter survival times even though cell number remained constant (Baldwin et al., 1997). Significant and similar differences in both hippocampal granule cell number and granule cell layer volume, but no differences in cell density, were found between superior and inferior learners among aged Wistar rat (Syková et al., 2002). Decreases in both hippocampal CA1 pyramidal cell number and pyramidal cell layer volume were also observed in monkeys after simian-immunodeficiency virus infection (Curtis et al., 2014). Unchanged cell numbers but decreased cell densities were found in adult human medullary nuclei when compared to infant ones (Porzionato et al., 2009). Hippocampal granule cell density was found to be highest in C57 mice when compared to DBA and NZB mice although total granule cell number in the three strains was the lowest in C57 mice (Abusaad et al., 1999). An increase in the density of cholinergic fibers and expansion of the width of the commissural-associational zone in the hippocampal dentate molecular layer after entorhinal cortex lesions were long interpreted as examples of reactive plasticity but later found to be secondary to molecular layer shrinkage (Phinney et al., 2004). Both primary visual cortex volume and the neuron number in schizophrenics were found to be lower than in controls but cell density did not differ (Dorph-Petersen et al., 2007). While vascular density increases in the cerebellum of Lurcher mice, total vascular length actually decreases (Kolinko et al., 2016). Even large increases in cell number can go hand in hand with decreases in cell density in the canary song system following androgen treatments (Yamamura et al., 2011). An age-related loss of hippocampal granule cells in APP/PS1KI mice was accompanied by an age-related increase in volume (Cotel et al., 2008; Cotel et al., 2012),

which would lead to a larger decline density than in number. Further examples from research on the morphological basis of neuropsychiatric disorders can be found in Dorph-Petersen and Lewis (2011). In each of these cases, conclusions drawn based on the numerator of a density obtained from representative sections would have been misleading because of changes in the denominator. Unknown changes in the denominator of a density have also been named the "reference trap" (Brændgaard and Gundersen, 1986). Changes in density indicate changes in the functional relations between the structures that provide numbers for the numerator and the denominator, and changes in density must have a cause. But the cause is not necessarily a change in the measure of one's primary interest (for a similar argument and further examples see, e.g., Mayhew, 1996).

5.2 A statistically representative sample of sections

One way to draw a representative sample from a series of sections of a brain region would be akin to drawing lottery tickets. Each lottery ticket has the same (a *uniform*) chance (*random*) of being drawn, and the chance that a ticket is being drawn is *independent* of the chance that another ticket is being drawn. This type of sampling is therefore referred to as *uniform random independent sampling*. Sections that are selected in this manner would constitute a statistically representative sample of the region that has been cut. With the exception of studies that look at the effects of sampling, this approach is hardly ever used in quantitative morphological studies. First, it is actually more tedious to draw a random sample than one may expect. Just fishing with a brush for a section in an Eppendorf tube is not good enough – large sections (or small ones – who knows) may be more likely to stick to the brush than small ones. One would have to collect sections in a way that would allow their selection according to some formal randomization procedure. The frequently used phrase "randomly selected" is hardly ever accompanied by a description how randomness was achieved. Second, it is counterintuitive and may be disruptive to other procedures. Randomization would mean that a sample from one animal actually may not contain any of those cherished anatomically typical sections, while the sections sampled in the next animal may contain all those cherished sections. Also, there is an intuitive resistance to the large variability that one correctly may expect across this type of samples.

Another way of representative sampling is much closer to procedures already in place in many laboratories – *uniform random* ***systematic*** *sampling*. We rarely collect all sections of

larger brain regions to use one particular stain on them. Instead, series of sections are collected in which the distance between the sections is determined by the needs that are perceived with regard to anatomical coverage (the series ought to contain examples of the typical appearance of the region of interest) and to other stains that we want to apply to the region. If four antibodies will be used, we need four series, maybe an additional series to try out antibody concentrations, and maybe one or two series in case something should go wrong or in case another stain is considered useful later in the course of the study. The four antibodies, trial sections and backup section require, in this example, a total of seven series to be cut. The seven series may then be collected into Eppendorf tubes or well plates, each tube or plate containing every seventh section that was cut. Each series is a *systematic sample* (every seventh) of the region of interest. To make these systematic samples statistically representative only *one small additional step* is required. Each section has to have the same (*uniform*) chance (*random*) to give their opinion with regard to the antibody used as any other section. We *cannot* always use series one (containing sections 1, 8, 15 …) for antibody A and always use series two (containing sections 2, 9, 16 …) for antibody B. If we did, sections 1, 8, 15 … would never be allowed to give their opinion on antibody B and sections 2, 9, 16 … would never be allowed to give their opinion on antibody A. Instead, we must pick one of our seven series at random when we assign them to a particular stain. This is the only step required to turn a traditional series of histological sections into a statistically representative *uniform random systematic sample* of sections.

In section 11 *Quantitative Morphology at the Bench* the number of series to be cut (any number is good, but some numbers give more options than others) and ways to deal with missing sections will be discussed.

5.3 Statistically representative sites within sections

Similar to the sampling of sections, sampling within sections must be statistically representative. Again, one has to resort to either uniform random independent sampling or to uniform random systematic sampling. Similar to the sampling of sections, uniformly and randomly sampling the area of the section in a systematic way requires us to randomly select a starting point and proceed from the starting point at regular intervals along the two dimensions of the area of the section. In the most common cases of square or rectangular

grids of sampling locations, the distances between sampling locations are often referred to as the x- and y-step sizes .

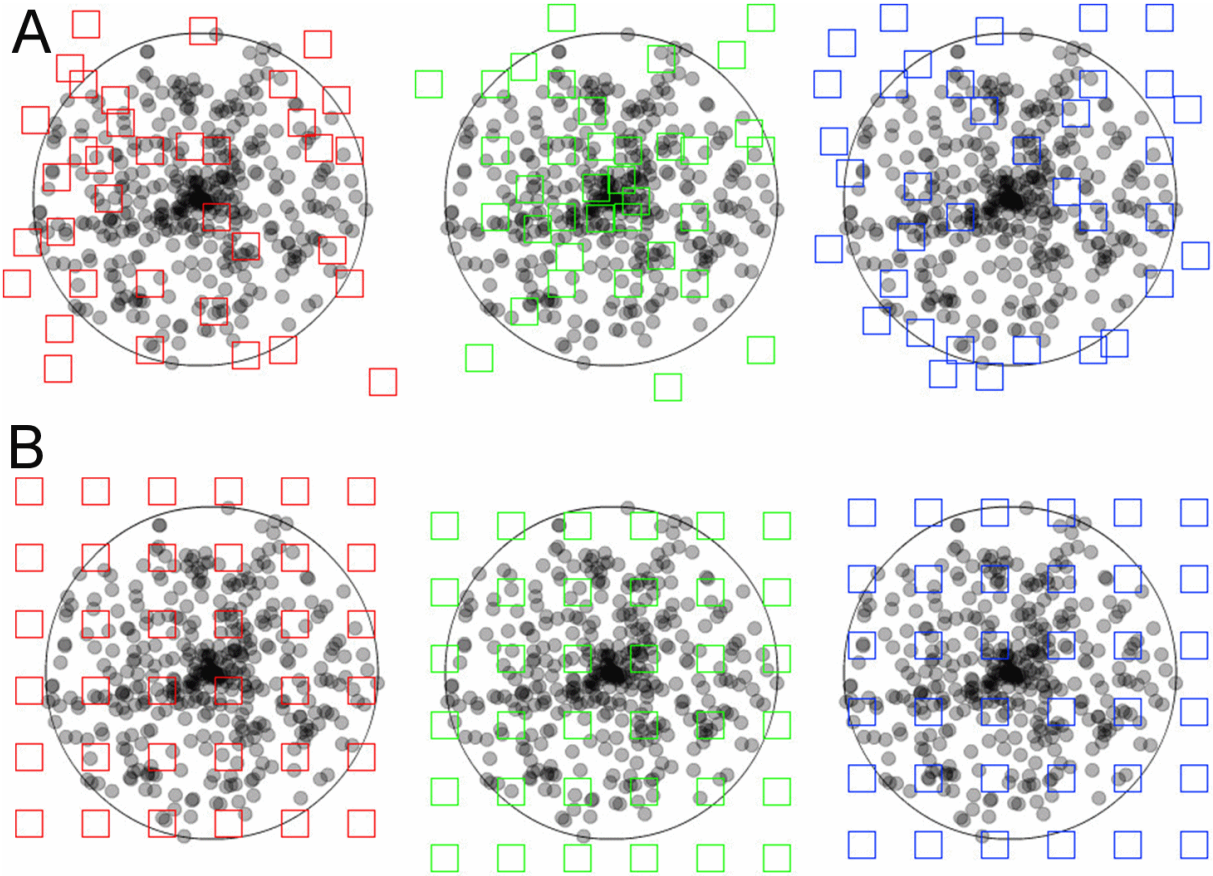

Figure 7 Uniform random independent sampling and uniform random systematic sampling. The area of a section containing unevenly distributed objects is probed in (**A**) three uniform random independent samples of the area (red, green and blue squares) and in (**B**) three uniform random systematic samples of the area. Both types of sampling are statistically representative. However, note that the systematic samples, unlike the blue independent sample, never completely miss the central part of the section, in which the objects of interest are spaced closely together. Also, they never, like the green independent sample, contain an unduly large number of samples in the central part of the section.

Both uniform random independent sampling and uniform random systematic sampling are illustrated in Figure 7. The distribution of dots in the circular region of interest may resemble the distribution of ganglion cells in the retina – they are spaced closer to each other in the region of the center of the visual field than in its periphery. Figure 7A illustrates three of the infinitely many possible uniform random independent samples; Figure 7B three of the

infinitely many possible uniform random systematic samples. Figure 7 illustrates a strength of systematic sampling. If the objects of interest are unevenly distributed within the region of interest, a systematic sample is more likely to capture this heterogeneity than an independent sample.

Just from the visual impression of the distribution of the sampling locations in the section, the red independent sample appears to probe areas of different density "just about right". The green independent sample seems to have "too many" sampling locations within the dense part of the section, whereas the blue independent sample seems to miss the dense part of the section almost completely. As long as these three possibilities are equally likely to occur, it does not matter. Across the sampling of several sections (or across several individuals), the differences between sections will average out to the correct group mean. In contrast to the independent samples, the systematic samples do not show a visually apparent overemphasis on parts of the region either densely or sparsely populated by the objects. The systematic placement of the sampling locations together with the distance between the sampling locations makes it impossible to happen. As a consequence, the variability of estimates obtained from a systematic sample may be lower than the variability obtained from an independent sample. Less variability between the samples typically also means less variability between subjects in a group and a greater chance to detect statistical differences between groups. In biological regions of interest, gradual changes in the density of the objects of interest are common, and the efficiency of systematic sampling can be expressed in simple mathematical terms. Variability between estimates decreases typically with a factor of $1/\sqrt{number\ of\ samples}$ when independent samples are used, whereas it decreases with a factor of $1/number\ of\ samples$ when systematic samples are used (Gundersen and Jensen, 1987; Roberts et al., 1993). That means that if a certain precision can be obtained analyzing 100 (or 64 or 16) samples that were collected in a uniform random independent manner, the same degree of precision may be obtained from only 10 (or 8 or 4) samples that were collected in a uniform random systematic manner. Importantly, it also means that it will typically require much less work to generate a precise outcome using a uniform random systematic sample.

Of course, Figure 7 has been drawn to make a point and may considered "unfair" to uniform random independent sampling with regard to the differences between the three samples. However, it is far from exaggerating what might happen when independent samples are

used. That all samples fall outside the region of interest and return a count of zero is statistically just as likely as all sampling locations hitting the central part of the section returning a count of very, very many. Uniform random independent sampling is playing a lottery of samples – any combination of sampling locations is just as likely to occur as any sequence of numbers in a lottery. Another example of the difference in the efficiency between independent and systematic samples is provided in section 7.2 *A point probe to estimate an area*.

The efficiency of systematic sampling, which was illustrated for the sampling within sections in Figure 7, applies to all levels of the sampling scheme. If the area of a region of interest shows gradual changes in size from section to sections, using a systematic sample of sections will not only be more conform to routine laboratory procedures but also more efficient than using an independent sample. Depending on the demands of the study, the sampling scheme may be extended to additional levels – like a sample of brain slices (Dorph-Petersen et al., 2009) from which a sample of blocks are prepared, which are then sectioned and, again, sampled (Lyck et al., 2009).

There are two special cases in which systematic samples do not compare favorably with independent samples. If the region of interest shows truly random fluctuations in size from section to section or if the objects of interest are distributed at random within the sections, the variability of estimates obtained from systematic or independent samples will be the same. The variability of estimates obtained from a systematic sample may be larger than that obtained from an independent sample if there are periodic changes in the size of the sections *and* a match with the intervals with which sections are drawn. The same is true for a match between the distances between sampling locations within a section and a regular periodic distribution of objects within the sections. The case of periodic anatomical change will be discussed in more detail in section 10.11 *CE estimators and systematic variations in morphology*.

5.4 Fractionator sampling

The *Fractionator* (Gundersen, 1986) allows calculating totals of number, length, surface or volume based on counts obtained from a sample of a region without any further knowledge about quantitative parameters of the region in which the counts were made. A uniform random systematic sample is taken at regular intervals, which allows calculating the fraction

of the region (containing the objects of interest) that is included in the sample. If every third section of the region was collected, the sample contains only one-third of the entire region and one-third of the objects that one may want to measure. The section sampling fraction, *ssf*, is one-third. If only part of the area of the section is investigated, e.g., one-tenth, only one-tenth of the objects of interest in the section will be contained in this sample of the area. The area sampling fraction, *asf*, is one-tenth. If one looks at the areas that were selected at high magnification, one may not look at every possible location along the thickness (z-axis) of the section but restrict analysis to, e.g., half of the thickness of the section. Again, only one half of the objects of interest that are located beneath the area will be contained in the sample. The thickness sampling fraction, *tsf*, is one-half.

Whatever we measure and however we perform measurements in the sample, we know how much of all-that-there-is we have looked at – one-half of the thickness of one-tenth of the area in one-third of the sections, i.e. one-sixtieth (1/2 × 1/10 × 1/3) of all-that-there-is. If what we measure is one-sixtieth of all-that-there-is, all-that-there-is in the entire structure must be sixty times what we measured. Uniform random systematic sampling and Fractionator sampling are two sides of the same coin. Uniform random systematic sampling becomes Fractionator sampling if we use the information about the sample to calculate the fraction of the region that we analyzed, and if we use this fraction to calculate the amount of all-that-there-is in the region.

The number of fractions that are included in a Fractionator sampling scheme can be extended according to the practical demands of a study. If, e.g., the human neocortex is the structure of interest, it may be divided in a number of smaller blocks that can be cut and stained following standard protocols (Lyck et al., 2009). Not all blocks need to be processed as long as the fraction of blocks that have been processed is known. Although practically not (yet) very useful in the neurosciences, the sections that are being used do not need to be parallel, equally thick or evenly spaced (Baddeley et al., 2006; Gundersen, 1986) as long as each section has the same chance to contribute to the sample as any other section and as long as it is known which fraction of all sections was sampled. Using a uniform random independent sample that represents a known fraction of all sections would also be a Fractionator sample. The same applies to the other levels at which one may want to sample.

Based on the general independence of Fractionator sampling from uniformly repeated positioning, sampling schemes like the *Smooth Fractionator* (Gardi et al., 2006; Gundersen,

2002) and the *Proportinator* (Gardi et al., 2008) have developed that can take into account regional differences in the distribution of the structures that we may want to know something about. For the same investment of work, they can return even more precise estimates than uniform random systematic samples.

5.5 No sampling

Correct sampling is important, but it is only a means to reduce workload and not an inevitable part of design-based stereology. Workload may not be prohibitive to the assessment of everything or, at least, everything at one particular level of the sampling scheme. If one can assess all sections but not all objects within them, one only needs to sample within sections. If there are too many sections but only few objects in each section, one only has to sample sections but not within sections. At each step at which sampling can be avoided, a source of variability can be avoided. A nice example of no sampling is the study of ganglion cell distribution in retinal whole-mounts by Coimbra et al. (2014). In that the retina can be prepared as a whole-mount, there is no need to section. With only one whole mount of each retina being available, sampling is not even possible. The section sampling fraction must be one. The depth of the entire retinal ganglion cell layer can be assessed with high magnification lenses, and it is technically not necessary to restrict sampling to part of the depth of the tissue. The thickness sampling fraction can therefore also be one. If workload is not a prohibitive factor to intensive or even exhaustive sampling at one or more levels of the sampling scheme, the question remains if the work is sensibly spent (Gundersen and Østerby, 1981). This question will be addressed in section 10 *Good enough? – Estimate Precision*.

6 A brief introduction to probes

Probes are the tools with which the amount of objects, length, surface or volume in a sample can be estimated. While sampling determines the place at which a measurement is being made, the probe that is selected determines *how* a measurement will be made. Stereological probes resemble other probes commonly used to investigate tissues. First, there is a similarity of the type of probe and the "thing" that is probed for. Proteins, in the form of antibodies, can be used to immunocytochemically probe for the proteins in tissues. *In situ* hybridization uses RNA probes to detect RNA in tissues. Not surprisingly, numbers of test points, lengths of test lines, surfaces of test areas and test volumes are used to probe for volume, area, length and number. Traditional and stereological probes share another feature – complementarity or the lock-and-key principle. Antibodies need to be matched to their antigens and RNA probes need to be complementary to the sequence that they are supposed to detect. There is a similar requirement relating stereological probes to the morphological parameter that they measure. If one is interested in the quantitative morphology of three-dimensional structures, *the dimension of the probe and the dimension of the parameter that is being measured must sum up to at least three*. A point (0-dimensional) can be used to estimate volumes (3-dimensional; 0 + 3 = 3); a line (1-dimensional) can be used to estimate areas (2-dimensional; 1 + 2 = 3); an area (2-dimensional) can be used to estimate length (1-dimensional; 2 + 1 = 3) and a volume (3-dimensional) must be used to estimate numbers (0-dimensional; 3 + 0 = 3). There is no proof of this concept, but, so far, no method has been found that would work despite the sum being smaller than three.

If the dimensions of probe and parameter do not fulfill this requirement, the probe will start cross-reacting with other parameters. This is akin to an antibody of insufficient specificity that cross-reacts with a protein different from the one it was intended to react with. The example in section 4 *The Problem* illustrated what happens if this requirement is not fulfilled. Not only do we generate the wrong number if we estimate number (0-dimensional) with a count in an area (2-dimensional; 0 + 2 = 2), but the wrong number depends on the size, shape or orientation of the objects that are being counted. A probe that we aimed at the number of objects cross-reacts with the size, shape or orientation of the objects.

The advantage of the dimensions of the probe and the parameter to sum up to three is that one can simply count the interactions between probe and parameter – the number of times that point probes fall within a volume, that a length pierces an area probe, that a line probe pierces a surface and that objects are contained within a volume probe (Figure 8). The counts and the size of the probes enter into, yet again, very simple equations that allow the calculation of densities. These equations are referred to as relationship equations.

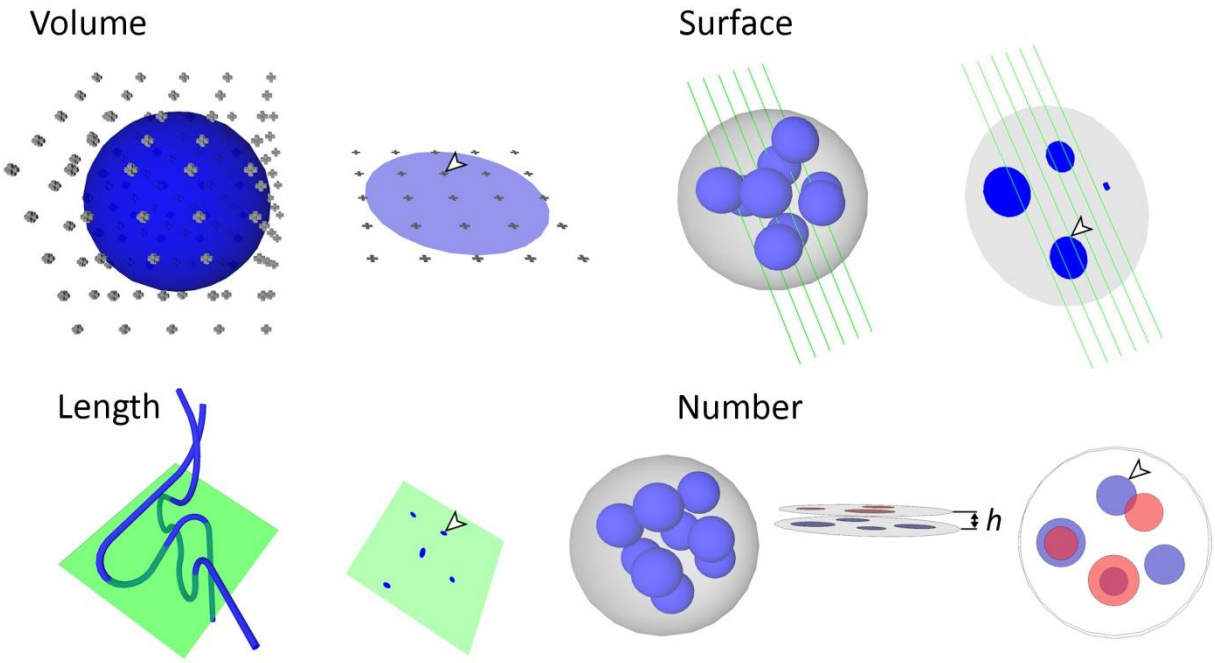

Figure 8 **Probes for *volume*, *surface*, *length* and *number***. Each probe is depicted how it would look like applied to the three dimensional tissue and how it would look like in sections that have been prepared from the tissue. For *surface* and *length*, the sections have been placed in the plane of the line and area probes. For *number*, the two sections and the distance (*h*) between them define the probe volume of the Disector. The two sections have also been superimposed on each other in the last view presented. For each probe, one of the places in which probe and feature interact have been marked by arrowheads in the sections. The total number of probe feature interactions in the examples are nine (*volume*), ten (*surface*), five (*length*) and two (*number*).

The relations ship equation for volume, length, surface and number densities are:

volume $$V_v = P_P = \frac{number\ of\ points\ counted}{number\ of\ test\ points\ that\ were\ applied}$$

(Glagolev, 1933, translated in Glagolev, 1955)

surface $$S_V = 2I_L = 2 \times \frac{number\ of\ intersections\ counted}{length\ of\ the\ test\ lines\ applied}$$

(Saltykov, 1946 as quoted in Saltykov, 1974; Smith and Guttman, 1953)

length $$L_V = 2Q_A = 2 \times \frac{number\ of\ profiles\ counted}{surface\ of\ the\ test\ area\ applied}$$

(Saltykov, 1946 as quoted in Saltykov, 1974; Smith and Guttman, 1953)

number $$N_V = Q^-{}_V = \frac{number\ of\ profiles\ observed\ in\ only\ one\ of\ two\ sections}{test\ volume\ \ encompassed\ by\ the\ two\ sections}$$

(Sterio, 1984)

Q refers to "Querschnitt", the German word for profile or cross-section. The superscript minus in number density refers to the "seen in one section but *not* the other". Reviews of the historical development of the relationship equations were presented by Hykšová et al. (2012) or, with an emphasis on mathematical theory, by Cruz-Orive (1997).

Next, densities can be converted to estimates of total number, length, area or volume using equally simple equations – *if* the structure of interest has been sampled correctly. The conversion of densities to totals will be addressed in the sections on specific probes.

In Figure 8, the probes for surface and length were placed in the plane of the section. This is how the probes would be intuitively applied, by defining an area of the section or by placing lines on the section and count intersections. However, a requirement for *all* probes is that *the number of the interactions of probes with volume, surface, length or number must only dependent on the amount of volume, surface length or number*. Section 4 *The Problem* showed how the Disector accomplished this for a number estimate. Section 9 *Length and surface estimators* will briefly describe why orientation may impact on estimates of surface and length and how tissue preparation or special shapes of test areas or lines rid these estimates from the influence of orientation. For estimating volume or number, there are no further theoretical requirements, but a number of practical constraints that are discussed together with the probes in the following sections.

The dimensions of the probe and of the parameter can sum up to more than three, and such probe/parameter combinations are part of some stereological methods. However, to obtain an estimate of the parameter does now require measurements instead of counts. If the size of a surface is, e.g., estimated by area probes, the length of the lines of intersection of surface and area probe need to be measured to obtain an estimate of the area. Depending

on the method used, the measurement may either not be trivial to perform or require special tissue preparation techniques. On method that uses line probes to estimate volume is the *Nucleator* (Gundersen, 1988), which has gained some popularity.

Lastly, if, e.g., cells are grown in a true “2-dimensional” monolayer cell culture, the sum of probe and feature only needs to be two ("Petri-metrics" in Howard and Reed, 2010). A test area is sufficient to count cells and lines may be sufficient to estimate the length of their processes. If cells could be grown in one line on a piece of “1-dimensional” string, probing a length of the string would be sufficient to estimate cell number. In these cases, some of the sampling strategies mention in the preceding section, or the test area described in section *8.2 The unbiased counting frame* will still be useful tools. If it turns out that these systems have three dimensions after all, e.g., if the volume of the cells in the monolayer cell culture needs to be known, we are back at a sum of three and the methods described in this thesis.

7 Probing Volume: the Cavalieri Estimator

The estimation of the volumes of brain compartments may serve different purposes. It is relatively easy to generate precise volume estimates. Volume estimates may therefore be an efficient first means to assess the likelihood of morphologically more specific structural changes. If changes in the numbers of neurons or synapses or dendrites or vessels are reflected in changes of the gross volume of the structure that contain them, the volume change may be easier to detect than the underlying more specific changes. If biological variability is low, differences as small as ~5% have been detected statistically in moderately sized groups (Slomianka and West, 1987) because rather little effort is required to increase the precision of the estimates. Also, volume estimates may be necessary if fractionator sampling (see section 5 *Sampling*) is not possible or desirable. In this case estimates can be generated from density estimates and reference volumes (Pakkenberg and Gundersen, 1997; West and Gundersen, 1990). Lastly, area estimates, which are part of the generation of volume estimates, are helpful in the design of sampling schemes that aim at other parameters than volume (see section 10.12 *Designing a useful sampling scheme*).

7.1 Calculating volume from area estimates

The mathematics that allows the calculation of volumes of a structure from the areas of parallel sections through a structure was brought into its final form by Bonaventura Francesco Cavalieri, a 17th century Italian mathematician. Following Cavalieri's theorem (Evans, 1917), the volume of a structure is equal to the sum of the areas of the structure in parallel sections that pass through it, multiplied by the distance between the subsequent sections, or

$$Volume = \Sigma Areas \times distance\ between\ areas$$

There is no more to it than that. We do know the distance between subsequent sections of our histological series. Note that the distance between areas is the distance between the same surfaces of the sections used in the series. If all sections are used there is no gap or distance between the sections, but still is a distance corresponding to section thickness between the top surfaces of two adjacent sections. In addition to distance, the only thing needed to calculate volumes are estimates of the areas that a structure of interest occupies in the sections. What comes to mind immediately is to outline the structure in some graphics

application and have the application calculate the area. As will be discussed in section 7.4 *Outlines or point counts?*, this may not be the most convenient way to estimate an area, and it is prone to errors. Instead, we can use point counts to estimate an area. But points (0-dimensional) and area (2-dimensional) only sum up to two? Correct, but here the sections are treated as 2-dimensional views of region of interest in which we are allowed to estimate area by points. We must do so because we cannot cut infinitely thin sections (see section 7.5 *A caveat: overprojection*).

7.2 A point probe to estimate an area

Imagine a region (blue circle in Figure 9) that occupies an unknown area of a reference area (all of the square in Figure 9). If a point is placed many times at a random position within the reference area, the number of times that the point will fall onto the region depends on the amount of the reference area that is occupied by the region. If the region would occupy all of the reference area, a point would fall onto the structure each time it is placed in the reference area. In this case, the probability of the point to fall onto the structure is 1. If the region would only occupy half of the reference area, a point would fall onto the region in only about half of the trials, i.e. the probability is about 0.5.

Usually it is the other way around – we do not know the area of the region, and we use the probability of randomly placed points to fall onto it to estimate the area. The more trials are made the better is the estimate of the area from the probability of points falling onto the region. The probability that we observe is equal to the proportion of the reference area occupied by the region – i.e. if we observe a probability of 0.48 (12 points out of 25 that hit the structure in Figure 9), about 48% of the reference area is occupied by the region. Note that this is the explanation for P_P in the relationship equation for volume density.

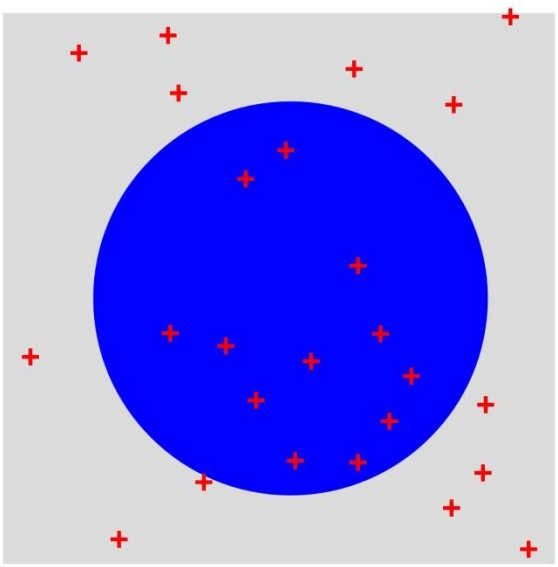

Figure 9 Using points to estimate an area. The reference area (entire area of the square) is probed with 25 randomly placed points to estimate the area of the blue circle within it. 12 of the 25 points fall onto the circle. The area of the circle can therefore be estimated to $12/25^{th}$ of the reference area. In that we arbitrarily can choose the size of the reference area, we can calculate an estimate of the size of the circle.

The area occupied by the region can now be calculated by multiplying the probability with the size of the reference area. We could arbitrarily decide on the size of the reference area before we start this little experiment, although, in the end, we will not have to do that.

In Figure 8 and Figure 9, the points are represented by a small cross hair, which is the most common representation of point probes in illustrations and also in stereology software packages. Unfortunately a cross-hair does not always allow deciding if the points hit the area of interest or not. First, it is not the entire cross that needs to fall onto the area of interest for the cross to be counted. Even the intersection of the bars of the cross does not always allow us to see if this point is located inside or outside the structure (Figure 10A).

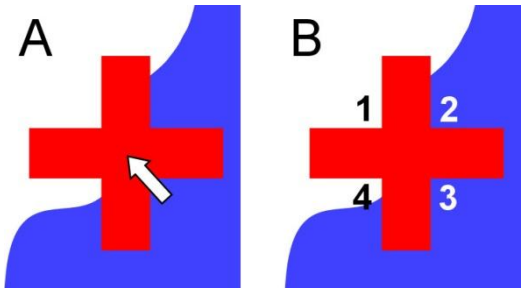

Figure 10 Representing a point probe. The representation of a point probe, a cross hair, falls onto the boundary of the blue region. In (**A**), the boundary is hidden by the cross hair, and it is not possible to decide if the intersection of the arms (arrow) falls onto the region or not. Using either one of the four corners of the cross hair in (**B**) allows an unequivocal decision. Corners 1 and 4 are falling outside the structure, while corners 2 and 3 fall inside.

The corners of the cross-hair provide better probes (Glagolev, 1955). In Figure 10B, two of the corners are located inside the structure while the two other corners are located outside the structure. Note that it is the very point at which the arms of the cross hair meet that is

used as a probe. Any one of the four corners can be used as a probe, but which one should be decided upon before the probe is applied to the section. In the survey of probes (Figure 8), the lower left corner was selected and generated a count of nine. The upper right corner would have generated a count of eleven.

A point-grid as an area probe – Using points that are placed completely at random within the reference area would represent a random independent sample of points. Following the uniform random systematic way of sampling, the area of a region present in a section can also be probed by placing a grid of regularly spaced points over the reference area (Figure 11). Each point of the grid is not only a probe for the entire area covered by the grid, but also for a smaller area associated with this point. If the points are, e.g., spaced 1 cm apart from each other, this smaller reference area is 1 cm^2. We can actually look at this grid in two ways. First, the 25 points may be understood as one uniform random systematic sample of points probing the entire reference area. In this case we still would need to know the reference area to estimate the area of the region. However we may also understand the 25 points as 25 samples that each probes the smaller areas once. Using this viewpoint, it is no longer necessary to know the size of the reference area.

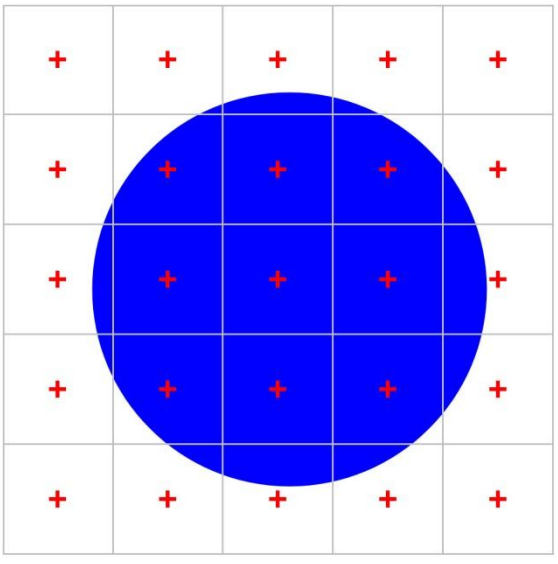

Figure 11 Using a point grid to estimate area. When the reference area is probed with a grid containing 25 points, nine points fall onto the blue circle. We estimate the area of the blue circle to $9/25^{th}$ of the reference area. Alternatively, we can look at this sample as 25 smaller areas, each with $1/25^{th}$ of the full area, that are each probed with one point. The area estimate of the blue circle would correspond to 9 times the smaller area, i.e. we do not need to know the reference area but only the area associated with each point.

Within each small square, the probability to hit the structure with the point will either be 1 (the point hit the region) or 0 (the point did not hit the region). In Figure 11, we obtain 9 probes of the smaller reference areas in which the probability of the point to fall onto the structure was observed to be 1 (1 hit/1 trial) and 16 tests in which this probability was 0 (0

hits/1 trial). An estimate of the area would be 9 × 1 × 1 cm^2 + 16 × 0 × 1 cm^2 = 9 cm^2. Increasing the size of the larger reference area (or the grid) would only increase the number of probes that return a zero probability, but it will not change the estimate of 9 cm^2. The area of the structure is directly proportional to the number of points that hit it. This number multiplied by the area associated with each point is an estimate of the area of the structure, or

$$Area_{region} = N_{points} \times Area_{point}$$

We do no longer need to know the size of the entire reference area covered by the point grid.

The precision of an area estimate – As already mentioned, the estimate of area would be more precise, if we repeated the random placement of the points many times. Instead of adding more points by repeating the estimate, we can use more points for each estimate, i.e. place the points closer together in the grid that we apply to the structure.

In Figure 12, points counted in the green squares return a determination of the area that is associated with the points. A hit is seen, the observed probability is 1, and the area is indeed one time the area associated with each point. The red area of the region is only estimated. Using the coarse point grid, 4 hits are seen in the 16 squares that contain some red area. For those 4 points we add the entire area associated with the point to our estimate even though the red area occupied less than that. Statistically, this overestimate will be balanced by the remaining 12 squares, which also contain a little bit of red area even though the points of these squares do not fall onto it and are not counted. Increasing point density four-fold in Figure 12, the green area, for which we obtain an exact area, increases. At the same time, the red area that is only estimated decreases. Not only does it decrease, it is also probed 26 times (each little square that contains a little red) instead of 16 times, which should provide a more precise estimate.

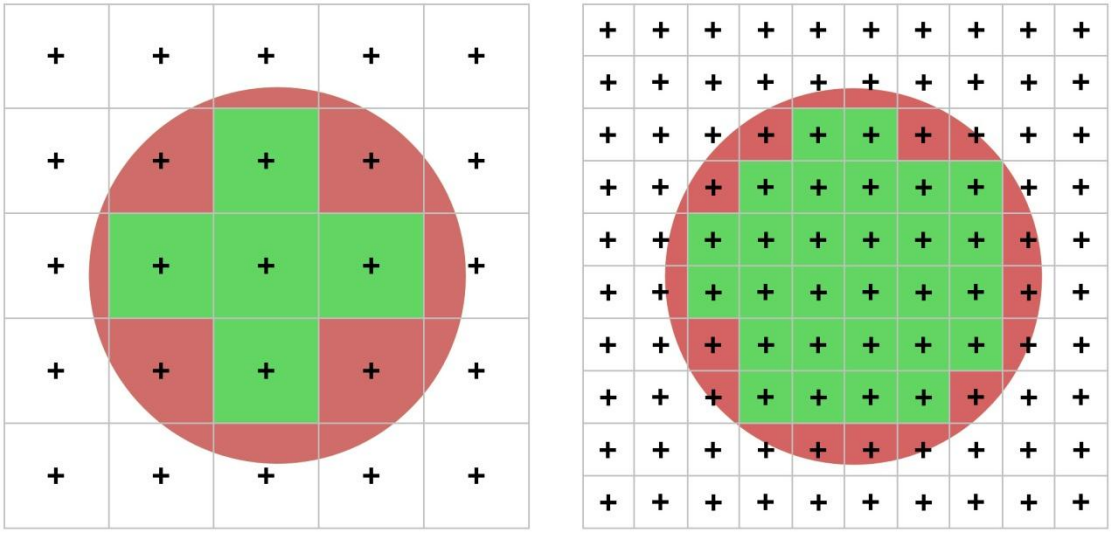

Figure 12 Increasing the density of a point grid increases precision. Point grids of different densities are placed over the round profile of a region. If point density increases, the number of areas that need to be estimated (red) increases (from 16 on the left to 26 on the right), while their total size and their size relative to areas that are determined by a point count (green) decreases. With more trials available to estimate the size of a relatively smaller area, the precision of the estimate of the area of the round structure increases with increasing point density.

Note that it is at least possible for all points to fall outside the structure or for all points to fall inside the structure if they are placed in the random independent manner that was used in Figure 9. This cannot happen for the grids (uniform random systematic samples) of points used in Figure 11 and Figure 12. That means that when estimates that are repeated, we would see fewer extreme values when we use a uniform random systematic sample. If we apply the point grids to different animals, we are also likely to see fewer extreme values because of the placement of the grids. The standard deviation would therefore be smaller, and we would have a better chance to observe a difference between two groups of measurements. For the same number of points used, a grid (uniform random systematic samples) is more efficient than randomly placed points to generate estimates that can be used to document changes that may occur in, e.g., an experiment.

How large is the region of interest illustrated in Figures 7, 9 and 10? By now we have three estimates – 48% ($12/25^{th}$) of the reference area from the independent sample in Figure 9, and 36% ($9/25^{th}$) and 40% ($40/100^{th}$ or $10/25^{th}$) from the two systematic samples in Figure 11 and Figure 12. All three estimates are statistically valid, but we cannot yet decide which

one comes closest to the true size of the region. Aside from efficiency, systematic samples do have another advantage over independent samples. We can *estimate* the margin of error based on the number of points that have hit the region of interest. For the 36% estimate the margin of error is ~ 10% of the estimate and, for the 40% estimate it is ~ 3% of the estimate. How this margin of error is estimated will be described in section *10.4 Variance originating from volume estimator point counts in sections*. Note (1) that the estimate using the dense grid is almost within the margin of error estimated for the loose grid, (2) that the reverse is not true, (3) that the estimate of the independent sample is outside the margin of error of both, and (4) that the margin of error indeed decreased to less than a third by increasing sampling density by a factor of 4. From (1) to (3), we may guess that the fine grid delivered the best estimate – confirming the expectation that the finest grid should return an estimate that is closest to the true area of the region. The region of interest in the figures actually occupies 40.5% of the reference area, which is well within the margin of error of the estimate using the fine grid.

7.3 An example of a volume estimate

A full example of a Cavalieri estimate is useful (1) to illustrate how much time it takes to generate an estimate once the material has been prepared, (2) to provide a small dataset that can be used in the following sections and (3) to show how the observer impacts on the estimate.

Figure 13 provides images of a hamster olfactory bulb taken in a horizontal series of every 24^{th} 20 µm thick, plastic (methacrylate) -embedded, and Nissl (Giemsa) -stained section. The first section is placed at random in the first sampling interval (240 µm after the first appearance of the olfactory bulb). We therefore have a uniform random systematic sample of sections. A counting grid, in which the points are 500 µm apart along the x-axis and 250 µm along the y-axis, is superimposed onto each of the images. The grid was positioned at random onto each section, and the sections are therefore probed at uniform random systematic sampling sites.

The region of interest has been selected to be the combined granule cell, internal plexiform and mitral cell layers (schematic in Figure 13). They are treated here as one structure, and any point falling onto anyone of these three layers should be counted. We do not need to keep track of which layer the counts came from because we are only interested in their

combined volume. To obtain the area estimates needed to estimate the volume, a point count is performed for each image. The upper right corners of the crosses were selected to represent the points, but any of the corners will do. To calculate a volume estimate, only the total number of points counted across all sections is needed. Counts are nevertheless recorded per section because they are needed to estimate the precision of the volume estimate (in section 10 *Good enough? – Estimate Precision*). Recording counts per section is also necessary when the anatomical distribution of the volume (or number, length or surface) has a scientific interest (e.g., Amrein et al., 2015; Buckmaster and Dudek, 1997; Chen and Buckmaster, 2005; Slomianka and West, 1987)

Note that the x- and y-distances between the points are not the same. The layers are mainly aligned with the x-axis, and decreasing the y-axis distance provides a better chance that they are hit reasonably frequently in each section instead of passing un-counted through the wider y-axis gaps.

Along the dorso-ventral axis (from 240 µm to 3140 µm), the following sequence of counts was obtained: 2, 20, 15, 12, 12, 6, and 3. The total number of points counted is 70. How large is the volume of the combined granule cell, internal plexiform and mitral cell layers? The area associated with each point is 250 × 500 µm = 125,000 μm^2. The section are 480 µm apart, i.e. each point is associated with a volume of 125,000 μm^2 × 480 µm = 60,000,000 μm^3. We counted 70 points and the volume is 70 × 60,000,000 μm^3 = 4,200,000,000 μm^3 = 4.2 mm^3.

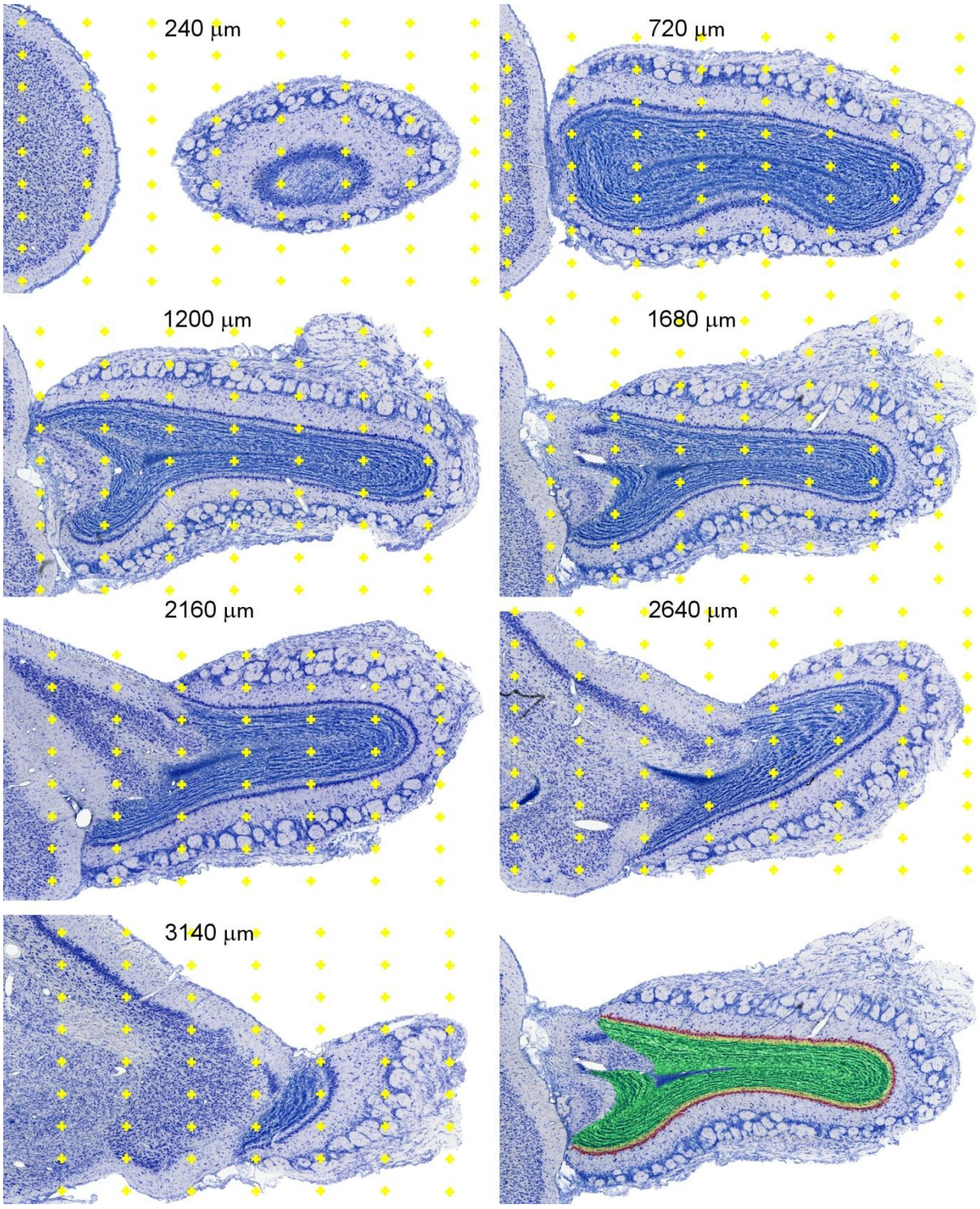

Figure 13 Cavalieri estimator applied to the hamster olfactory bulb. Nissl-stained, 20 µm thick plastic sections. The region of interest is highlighted in the lower right image: granule cell layer – green, internal plexiform layer – yellow and mitral cell layer – red. Their combined volume is estimated. Any point falling on any one of them is counted. Sections are 480 µm apart from each other. Distances between the points in the grids are 250 µm along the x-axis and 500 µm along the y-axes.

The counts quoted above and those of other observers are likely to vary a little even though the same images and points are used. This may be because of the use of a different corner. It may also be because of a different "translation" of the schematic, which was provided for only one section, to the other sections of the series. It may also be because criteria to count a point (How does the tissue in the corner of the cross need to look like to be counted?) vary from observer to observer. The initial term for design-based stereological methods – *unbiased* stereological methods – was, with some justice, criticized because of the two latter points. The Cavalieri estimator cannot generate a bias in its own right, but the data that it is fed with may vary with the interpretation of the material by the observer, i.e. with observer bias. Observer bias cannot be avoided by any method involving an observer. It even cannot be avoided by many methods that do not involve an observer, e.g., an automated, image analysis based assessment. The use of automated methods only transfers the observer bias to the person that at some time in the past calibrated the automated assessment. At least someone else could be blamed for mistakes.

7.4 Outlines or point counts?

Areas are often measured by outlining a structure and the subsequent automated calculation of the area of the structure based on the outline. How well the calculated area corresponds to the actual area does, of course, depend on how well the outline is defined. The complexity of the shapes and, therefore, the effort to define precise outlines may vary with the region of interest. In Figure 13, the region of interest was chosen to be the combined granule cell, internal plexiform and mitral cell layers because the border are rather clearly defined, which suits an exercise and also makes outlining rather easy. For the two thinner layers, the internal plexiform layer and the mitral cell layer, outlining becomes tedious work. An error that has only a small effect on the combined volume of the three layers will have a much larger effect on the volumes of the thinner layers. One may have to go to higher magnification to generate better estimate for the thinner layers – only to realize that borders which look well defined a low magnifications often present increasingly complex outlines at higher magnifications. Essentially, it is the observer's decision to accept possible errors that are associated with the fact that the precise outline simply cannot be seen or with the fact that the outline needs to be approximated because it is too complex to be traced precisely with a reasonable effort. Point counting does not have this problem. The

only decision that has to be made is whether the point falls onto the structure of interest or not. The number of times that this decision will have to be made depends on the point density and the area of the region, but it does not depend on the complexity of the outline of the region of interest.

Using outlines, an additional source of error may be associated with the calculation of areas based on the points that have been placed to define the outline. The area may, e.g., be calculated for the polygon that is defined by the points (Figure 14A) or based on smooth lines (splines) that are approximated to the points (Figure 14B and C). How well the area is estimated depends on how well the area fulfills the underlying assumptions of being a polygon or an area with smooth outlines. Although the resulting bias may become small with an increase in the number of points used to define the outline, it is not possible to guess at the resulting error without beforehand knowledge of the area.

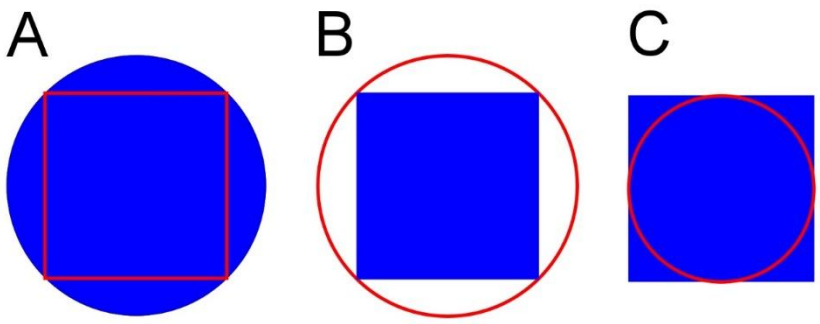

Figure 14 Polygonal- and spline-based approximations of an area to boundary points. (**A**) If a region is not polygonal, an area estimate (red) based on outlines will underestimate the actual area (blue) if the calculation is based on the area of a polygon. (**B** and **C**) If the area is estimated using smooth curves approximated to the points (splines), the area is, depending on the placement of points, over- (**B**) or underestimated (**C**) if the region is polygonal. The direction and size of the error depends on the number of points used to define the outline and the shape of the region outlined.

7.5 A caveat: overprojection

To estimate area, we had to treat sections as 2-dimensional images of the region of interest. Cavalieri's principle is valid if the sections of the structure are infinitely thin, but this requirement cannot be fulfilled; our sections do have a thickness. As soon as the "sum-up-to-three" rule is violated, cross-reactivity sneaks in. In this case, volume may be influenced by the 3-dimensional shape of the region of interest (Sterio, 1984). The conflict of theory

with practical constraints may result in an overestimate of volume – regardless if we are using point counts or outlines. Figure 15 illustrates the problem. The wish for thin sections when using the Cavalieri estimator of volume is in conflict with the wish for thick sections that almost all other quantitative methods will have. In histological practice (see below), it may be of little importance.

For a convex structure (e.g., a sphere), the error amounts to the largest volume that we estimate in a single section (Gundersen, 1986), and this volume could be subtracted from the estimate of the total volume of the structure. If the shape has dents or contains hollows, the size of the error will be larger.

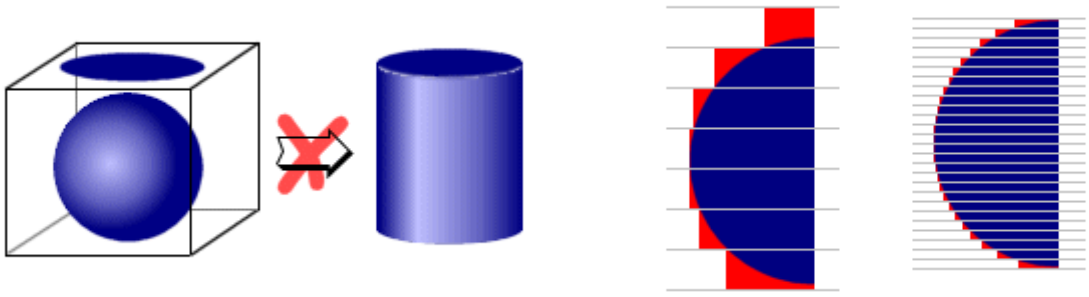

Figure 15 The overprojection error of volume estimates. If a section that is thick when compared to the region – here, a worst case in which it is actually thicker than the region – the volume of the region is overestimated. The area of the region will look like a circle, which, when multiplied with the thickness of the section, will give us a volume estimate of a cylinder rather of the sphere that the region actually looks like. Looking at sections of the sphere from the side, the error (red) decreases as sections become thinner and thinner. The errors made in each section will sum up to the volume of the largest of the sections.

Note that the error occurs when the border of the area that we see in a section is sharp and unequivocal even a low magnification – essentially like it would appear when a structure is opaque like the sphere in Figure 15. In analysis of MRI scans, a dependence of volume estimates on slice thickness has been noted that shows overprojection effect (Bonilha et al., 2003). In histological sections this will rarely be the case. The whole purpose of sectioning is to make a structure transparent so that we can examine its internal structure – to see the cells, their processes, vessels etc. Because of that, one is rarely faced with the problem of sharp boundaries, but rather with the problem of finding boundaries in the first place. How

much of the overprojection error remains under these circumstances will depend on the decisions of the observer. The size of the error will also depend on how many sections can be cut from the structure of interest. If hundreds of sections can be cut, even the largest section may only represent a percent fraction of the total volume. If only very few sections can be cut, one may decide to perform a point count at high magnifications. Because the thickness of the focal plane is usually thinner than the thickness of the section (about 0.5 µm for an oil-immersion objective), the size of the error would diminish to the volume contained in the largest focal plane.

7.6 Simple Cavalieri estimator implementation

The Cavalieri-Estimator is probably the method easiest and least expensive to implement. Several plugins have been implemented to generate point grids in ImageJ. Point grids can be prepared using any type of graphics software that by way of grids allow a placement of the test points with reasonably consistent point-to-point distances. Even recreational graphics software often contains a useable function, one of which was used to generate Figure 13. The point grids can next be placed over images of the region of interest – either digitally or, copied onto transparency sheets, onto prints. Transparencies can also be taped to a screen if a simple digital camera setup is available. The point grid can be calibrated with images of an object micrometer taken under the same conditions as the tissue. If a region is too large to be imaged on the screen or on a print, simple tools (Adiguzel et al., 2003; Kaplan et al., 2001; Melvin et al., 2007) allow regions to be probed across several fields of view of the microscope. In a nutshell, two (or more) points that are separated by a known distance along the x- and y-axes are marked on the screen. Once a field of view has been examined, a feature visible close to one of the points is moved in the x- or y- direction to the next point. Calculations of the volume estimates require no more than, in the lowest-tech version, paper and pencil.

8 Probing Number: the Disector

We have already probed for number using the Disector in section *4.2 An almost as easy solution*. The example used in Figure 5 and Figure 6 were based on the comparison of two real, *physical* sections. The variation of the Disector using this approach is referred to as *physical Disector*. Using this approach and the statistical demand that each object of interest, e.g., cells, must have a chance to contribute to the sample requires that it must be possible to decide for *every cell* if it is present in the counting section, look-up section or both. Although image analysis features of stereological software packages may facilitate the process, both the generation of material of a quality that permits the analysis and the analysis itself remain demanding. Consequently, few studies have been published that used the physical Disector (e.g., Korbo et al., 1990; Miki et al., 2003; Miki et al., 2004; West et al., 1988). Today, the physical Disector is used almost only if it is not possible to use the *optical Disector*, e.g., in electron microscopical quantitative morphological studies (see section *11 Quantitative Morphology at the Bench*). If the physical Disector is used, the two sections need to be assessed independently from each other (Hedreen, 1998a), i.e. the structures of interest are identified in each section and the two sections are compared afterwards. Two conceptual/technical advances, beyond the way in which the Disector was introduced in section 4 *The Problem*, have allowed its routine application in a large number of studies. The first, the unbiased counting frame (Gundersen, 1977), allows restricting the sample to a small fraction of the area of the sections. The second, the optical sectioning of thick tissue section (Gundersen, 1986; West et al., 1991) has made it unnecessary to compare two physical sections.

8.1 Which object should be counted?

In some tissues, e.g. blood, the answer to the question is obvious. If we are interested in the number of cells, we count cells. This answer is valid if a cell is easy to recognize. This is unlikely to be the case in the nervous system. Could we count a neuron when we for the first time recognize one of its dendrites? If not impossible, it at least seems very impractical. Even the soma may require closer scrutiny – is it actually two small somata we count, or is it the trunks of two primary dendrites that merge into a larger soma outside the optical plane we are looking at? One suitable unit to count would be the nucleus, which, in the nervous

system, typically has a fairly simple shape. Note that "simple shape" is not a requirement to estimate the accurate number, but that it simply makes it more convenient to estimate the accurate number.

What is more important than convenience is that the nucleus also has a fixed numerical relationship to the objects that we are actually interested in – there is one nucleus in every cell. The frequencies of binucleate neurons or glia in the scarce CNS studies that report them (e.g., Ackman et al., 2006; Rezze, 1966) are so low that they can be safely ignored in the vast majority of studies, but this may be different for the PNS (Forsman et al., 1989; Ribeiro, 2006). Another convenient unit may be nucleoli, which in some (but not all) cell populations and species have a 1:1 relation to the object of interest. In general, the key to the suitability of any structure to serve as a representative of the objects that we are interested in is a known numerical relationship between the structure we can count and the objects we want to count. This relation does not need to be 1:1, but it needs to be known. Most humans could be counted by counting hands (2:1) or fingers (10:1) but not by counting hairs.

8.2 The unbiased counting frame

Even a single section will often contain too many of the objects of interest to count them all. A smaller sample needs to be drawn in a way that provides each object with the same chance to contribute to the sample. The counting frame devised by Gundersen (1977) allows just that (Figure 16). The counting rule of the Disector allowed the unique identification of an object along the z-axis. In essence, the counting rules associated with this counting frame apply the same rules to the x- and y-axis of the section. One may imagine the x- and y-axis as "new z-axes" along which the section is "re-sectioned" (Figure 16B and C). The first cuts (parallel to the y-axis) divide the section in a series of stripes (Figure 16B). Applying the rules for the z-axis now along the x- axis, we are allowed to count what is seen in a stripe if it was not visible in the previous stripe. The next cuts (parallel to the x-axis) further divide the stripes into small rectangles (Figure 16C). Applying the counting rules now along the y-axis, we are allowed to count what is seen in a rectangle if it was not visible in the previous rectangle. Two sides of the rectangle end up becoming *exclusion lines*. If an object crosses these lines, it will also be present in one of the adjacent rectangles, in which it will not cross the exclusion lines and in which it would be counted. The borders of the rectangles that an object may cross and still be counted are referred to as the *inclusion lines*. These rules are

not quite sufficient yet. One object in Figure 16C does not cross an exclusion line, but it crosses the inclusion lines of *two* rectangles that touch each other with their corners. This object would be double-counted if we would look at all frames. Analog to the example in section 4 *The Problem*, spacing the frames far enough apart to avoid double-counting does not solve the problem. Instead, the exclusion lines are extended to avoid this from happening. If their length is at least as great as the largest diameter of the objects that are to be counted, no object has a chance to be double-counted (Figure 16D).

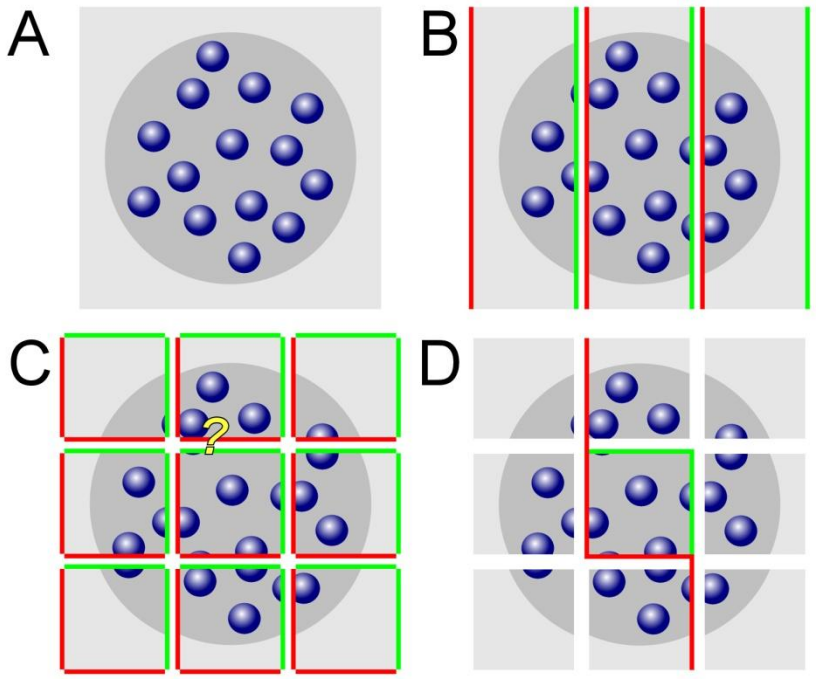

Figure 16 The unbiased counting frame. A section containing 14 objects (**A**) is successively divided first parallel to the y-axis (**B**) and then parallel to the x-axis (**C**). Along the division lines the same counting rules that have been established along the z-axis are now applied along the x- (**B**) and then x- and y-axes (**D**). 15 objects instead of 14 are detected if all objects fully contained within the squares and all object fragments touching a green inclusion line are counted in (**C**). The overcount is due to the double-count of the cell labeled with the question mark. The double-count is avoided by the extension of the red exclusion lines in (**D**).

Figure 16D represents the traditional, square counting frame with exclusion lines extending along the y-axis of the section plane. The counting frame may also be rectangular, and the exclusion lines may just as well extend along the x-axis as along the y-axis. As long as a complete tessellation, i.e. a complete covering by non-overlapping tiles of the surface of the section is possible (Figure 17), counting frames/exclusion lines may take on complex shapes

(Gundersen, 1977) – although, in practice, squares and rectangles are almost exclusively used.

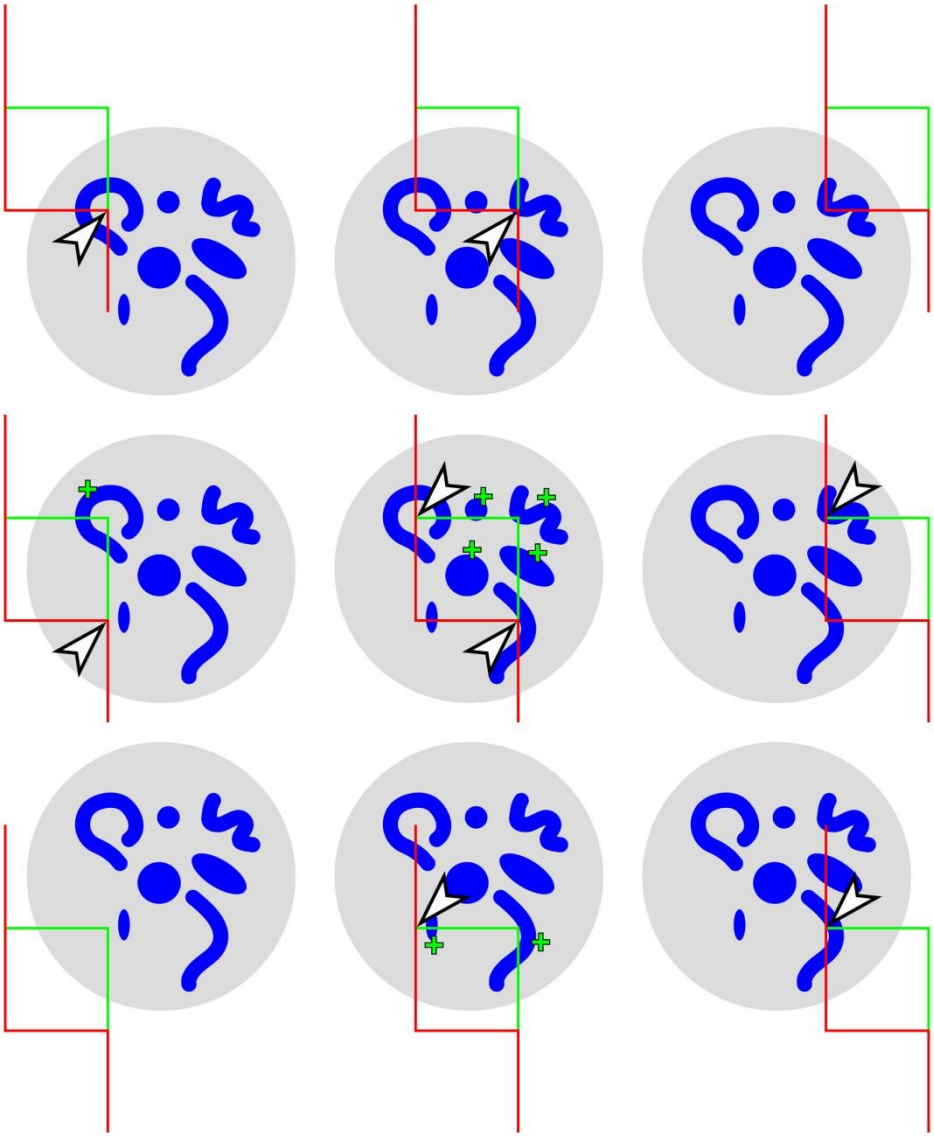

Figure 17 A complete tessellation of a region containing objects of varying size and shape. If the counting rules are applied, each object is counted at least once and never more than once independent of its size, shape or orientation. The objects are marked with a plus-sign in the image of the frame in which they are counted. Using some approaches, the number of counting frames applied to a region needs to be known. Sometimes one needs to know how many probes have been applied. Any selection of the corners of a counting frame, in this case the top-left and bottom-right corner, can be used to estimate the number of *entire* counting frames that fell on the structure. Eight of the selected corners (marked by an arrow) fall inside the grey region of interest. Objects were therefore counted in an area corresponding to 8/2 = 4 full counting frames.

In the description, “to cross” an exclusion or inclusion line was used as counting criterion. One may equally well use “to touch” as the criterion for inclusion or exclusion. Critical is the use of the same criterion for both the inclusion and exclusion lines of the counting frame. In the complete tessellation shown in Figure 17, each object will be counted once independent of its size, shape or orientation.

8.3 Optically sectioning a section

The acceptance that the Disector has found relates to the fact that its second incarnation, the *optical Disector*, is practically far less demanding than the physical Disector. Instead of comparing two physical sections, we count in optical sections, i.e., within a part of a physical section defined by two focal planes placed at different positions along the z-axis of the physical section. The counting criterion of the physical Disector, “not present in the look-up section but present in the counting section”, is simply exchanged by the corresponding optical criterion (Figure 18).

> *An object is counted if it is not recognized in top focal plane of the optical section, but first recognized as we gradually move, within the section, to a bottom focal plane of the optical section.*

The distance between the two focal planes, *h*, defines the height of the optical Disector in which we perform the count.

Typically, one would first recognize an object when the object appears in focus for the first time. Unfortunately, the perception of “in focus” may vary from observer to observer. One of the strength of the optical Disector is that it is not crucial to agree on a specific focal plane as long as each investigator recognizes each object to be counted at least once. This is because we use two focal planes to decide if objects should be counted or not. An investigator (early recognizer in Figure 18) may decide that, according to her or his criteria, the top-most object is in focus in the top focal plane that defines the the top of the optical Disector. Therefore, it cannot be counted. An object placed similarly in relation to the bottom focal plane will be counted. In this case, both decisions are formally correct. The top object was indeed already present above the top surface of the optical Disector and the bottom object appeared for the first time in the optical Disector. Another investigator (late recognizer in Figure 18) may only recognize the object when the focus is moved slightly

further into the section. The top-most object that was not counted by the early recognizer is counted (formally erroneously) by the late recognizer. If, however, this personal perception of "focus" is applied consistently to the top and bottom planes, the late recognizer will not count the object crossing the bottom focal plane of the optical Disector. The error made by the late recognizer close to the top plane has been cancelled by a second error made close to the bottom plane. Formally, the wrong objects have been counted, but in practice, the count is the same. Not every location will harbor objects that cancel possible errors caused by an individual perception of focus. Across all probe locations however, we should find as many objects in the upper position as in lower position in the many samples that should be obtained from the region that contains the objects. Errors will cancel even when someone recognizes the object while they are still out-of-focus if there is sufficient space below the optical Disector to compensate for the error. As far as the optical Disector is concerned, the focus of a student is likely to be as good as the focus of a professor.

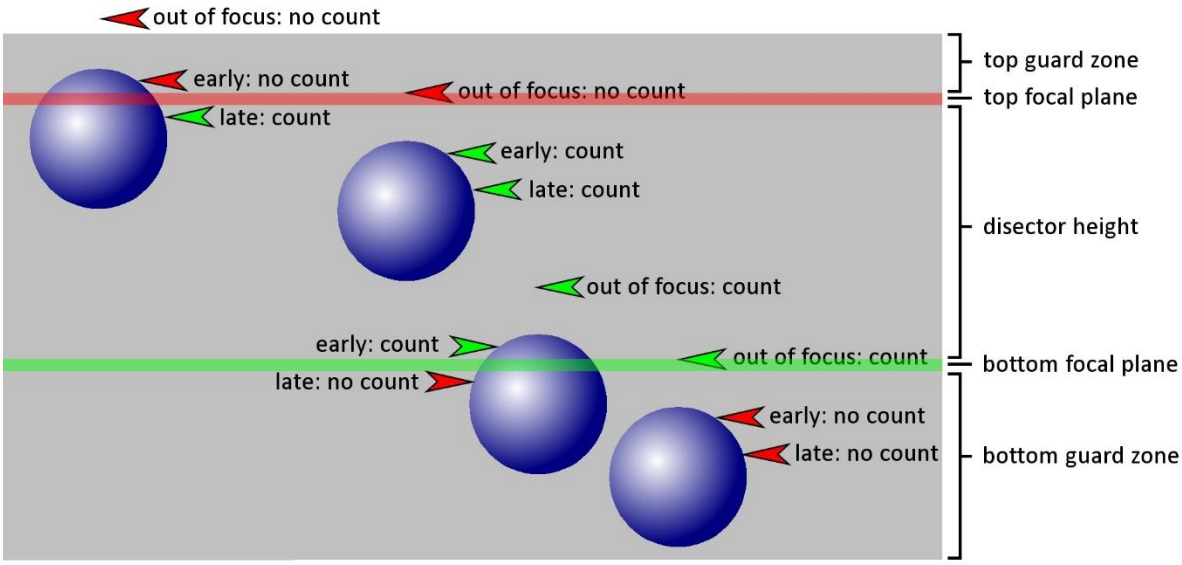

Figure 18 Independence of optical Disector counts from object recognition criteria. Three observers are looking at the same sample of cells. One of them (late) recognizes the object rather late, one of them (early) recognizes them rather early and one of them (out of focus) even recognizes the objects before they really comes into focus. Differences in the first recognition (arrows) of objects between observers compensate across the top and bottom focal planes that define the height of the optical Disector. Regardless how objects are recognized, each observer will count two objects (green arrows) in this example. Recognition will typically be defined as the object appearing in focus, but this is not a requirement of the probe. (adapted from Figure 7 in West et al. (1991))

Critical aspects of the application of the optical Disector are the ability to measure the distance between the top and bottom focal planes and to observe if an object can be recognized or not as we move the focal plane. Both demands are satisfied by the use of high magnification oil-immersion lenses with large numerical apertures in conjunction with a microcator – a device that measures the distance that the section moves along the z-axis as we focus through the section. It must be pointed out here that the physical distance which a section moves along the z-axis as the focus knob of the microscope is turned does not necessarily correspond to the distance that the focal plane moves within the section (Galbraith, 1955). This is only the case if the refractive index of the embedded tissue corresponds to the refractive index of the medium between section and microscope lens. For most embedding media this will be the case if oil-immersion lenses are used. If air or water-immersion lenses are used instead, the physical movement of the section along the z-axis will typically no longer correspond to the distance moved by the focal plane. If this situation cannot be avoided, conversion factors need to be applied – a task that can be automated by stereology software packages.

Figure 19 illustrates two optical Disector samples – both are fluorescence stacks of mouse hippocampal CA1 pyramidal cell nuclei recorded with a conventional, non-confocal microscope and either traditional epifluorescence illumination or structured illumination (Karadaglić and Wilson, 2008; Mertz, 2011; Neil et al., 1997). The latter is visually similar to a confocal stack. Although the visual appearance of the stacks is strikingly different, nuclei can be reliably counted in both of them. It appears that individual nuclei cannot be recognized in as many optical planes in the traditional stack when compared to the confocal-like stack. It does not matter as long as they can be recognized at least once. It also appears that they need to be a little larger before they are recognized in the traditional stack. It does not matter either because the position at which a structure is recognized does not matter as long as it is similar at the top and bottom planes of the stack.

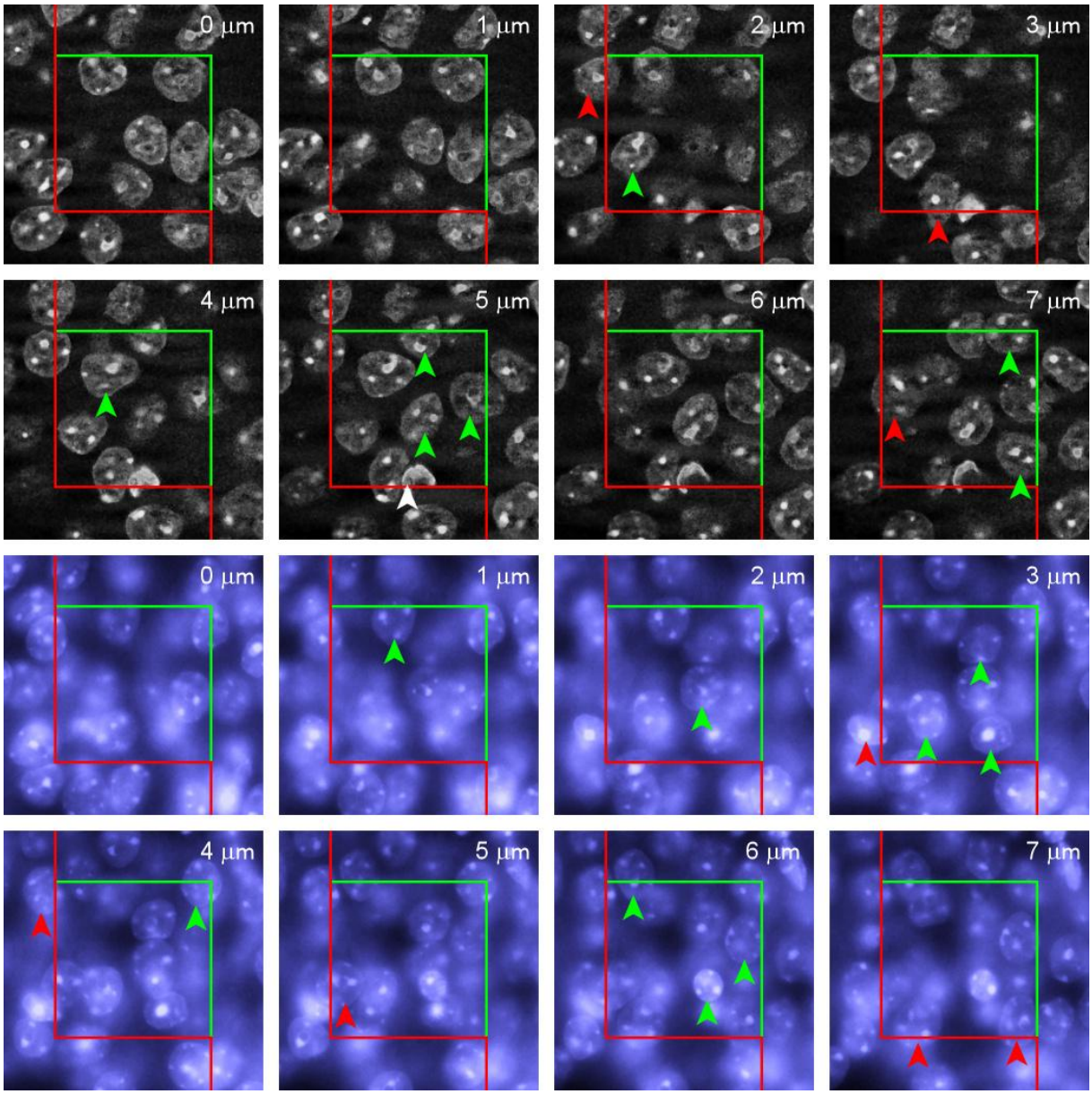

Figure 19 Examples of optical Disector probes. Image stacks representing two optical Disector probes that are 7 µm deep in a DAPI-stained mouse hippocampal CA1 pyramidal cell layer. The upper stack was recorded using structured illumination, which generates confocal-like images with little out-of-focus light. The lower stack represents normal epifluorescence images. In both stacks, nuclei marked with green arrow can be counted according to Disector counting rules, i.e. they are not present in the top focal plane of the Disector (0 µm), they do not cross the red lines at the depth at which they are first recognized (nuclei marked with red arrows), and they are contained within the counting frame or cross the green lines at the depth at which they are first recognized. The personal criterion for the recognition of a nucleus was a nice, crisp outline. The counting frame measures 30 × 30 µm. The white arrow at a depth of 5 µm in the first stack most likely represents a capillary endothelial cell nucleus that is not being counted.

8.4 Guard zones

This expression is used for the parts of the physical section located above and below an optical Disector. Their use may be recommended for two reasons. First, the surfaces of the sections may be "disturbed" by the cutting, meaning that objects that are cut or even just close to the surface of the sections may become distorted or damaged (see Helander, 1983 for an illustration of disturbances in paraffin and plastic sections). If the disturbance is great, one may not be able to recognize the objects that are affected (Andersen and Gundersen, 1999). If they cannot be recognized, they cannot be counted. This violates one firm requirement of design-based methods – namely that each object must be able to contribute to the sample. How far an optical Disector sample should stay away from the surfaces of the section depends on the depth of the disturbance and the sensitivity of the object to it.

Formally, it is best determined by a small pilot study in which Disector samples spanning the entire thickness of the section are used. While counting, the position of objects along the z-axis is recorded. After some objects have been counted, the numbers of objects located in intervals along the z-axis is plotted (Figure 20). One would not expect that the number of objects varies along the z-axis of the section, and the graph should ideally look like a rectangle. If objects at or close to the surfaces of the sections cannot be recognized, the numbers in the intervals at or close to the section surfaces will be smaller than in intervals located in the middle of the section. The Disector samples should only span the part along the z-axis in which the plot is close to horizontal. The remainder of the depth of the section should become part of the guard zones.

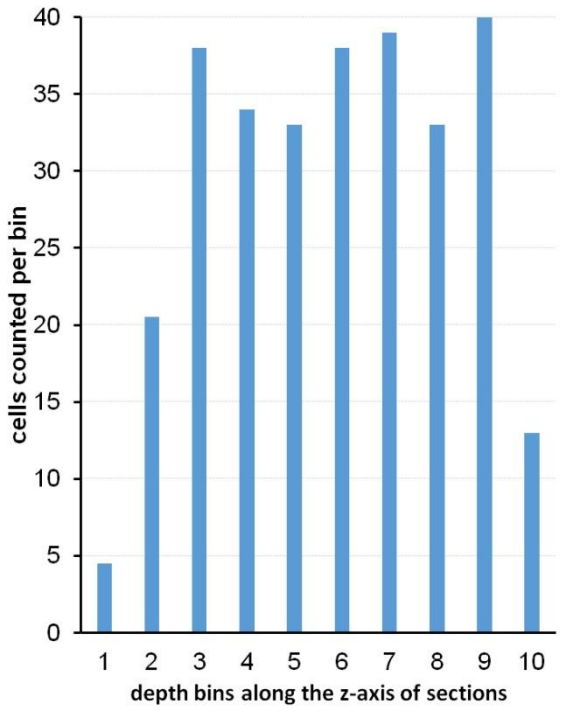

Figure 20 Z-axis plot of cells. Hippocampal CA3 pyramidal cells are observed throughout the depth of methacrylate sections. The depth of their first recognition is recorded. Cells already visible in the top focal plane of the section are omitted. If cells distribute evenly throughout the thickness of the sections, similar numbers should appear in each bin. In this example, cells distribute evenly from bin 3 to bin 9. Low numbers in bins 1 and 2 were due to low counts in the top 2 µm of the sections. Low numbers in bin 10 were due to low cell numbers in the bottommost 1 µm of the sections. Data were kindly collected by R.M. van Dijk in sections prepared for West et al. (1991).

The formal assessment of the distribution of cells along the z-axis presented in Figure 20 may be useful for another reason. It allows detecting if, e.g., an antibody used to stain the objects of interest has evenly penetrated the section. If that is not the case, one may observe a distribution along the z-axis that shows a depression in the middle of the section (Torres et al., 2006). This issue, and others that relate more to practical issue in the laboratory than to the principles of the method, will be discussed again in *section 11.4 Cutting, staining and coverslipping*.

The second reason why a guard zone may be useful is the shape of the objects that are to be counted. Some shapes may not allow to decide if a profile observed in the last focal plane of a Disector should be counted or not (Figure 21). If that is the case, the part below the last focal plane needs to be examined to make the decision. This part, the lower guard zone, should be large enough to allow deciding if a profile should be counted or not.

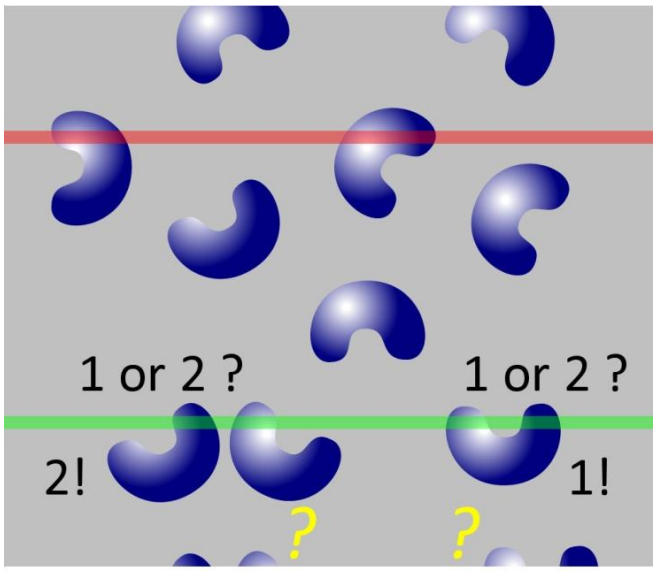

Figure 21 Guard zone and object shape. A deep guard zone may allow the correct identification of profiles if the shape of the objects is complex. In this example, the part of the section below the Disector needs to be examined to determine if object profiles visible in the lowest focal plane (green line) should be counted as one or two objects. This decision is not possible if the Disector would be extended to the bottom of the section.

Lastly, the guard zone below the Disector provides the space that may be necessary for objects to be counted using individually variable criteria for the recognition of objects. If objects are, e.g., recognized 5 µm before they actually are located in the focal plane, we may need 5 µm below the Disector in which cells can be counted to compensate for the error.

8.5 From Disector counts to total number: Fractionator and $N_V \times V_{Ref}$

Ironically, given the complaints about the use of density changes as evidence for changes in number, all that the Disector probes initially deliver is a number density estimate, N_V, i.e. the number of cells counted in the volume of all the Disector probes that were used to count. One could, of course, also calculate the average number per volume of one Disector probe or a number per a "unit of choice" of volume. There are two ways to convert the count to estimates to total number – the Optical Fractionator and the $N_V \times V_{Ref}$ method.

The optical Fractionator (West et al., 1991) – If uniform random systematic sampling was used throughout, we can safely ignore the volume information provided by the Disector. Uniform random systematic sampling also means Fractionator sampling (see section 5.4 *Fractionator sampling*). It is therefore known which fraction of the volume of the region of interest has been included in the sample. Total number can be calculated by multiplying the number of cells counted with the Disector probes with the inverse of the sampling fractions that have been used to sample.

$$N = \sum Q^{-} \times \frac{1}{tsf} \times \frac{1}{asf} \times \frac{1}{ssf}$$

If, for example, the height, *h*, of the Disector probes spanned 10 µm along the z-axis and section thickness, *t*, is 20 µm, the thickness sampling fraction (*tsf*) would be 0.5. The inverse of the thickness sampling fraction is 2. Note that section thickness here refers to the actual thickness of the stained and mounted section on the slide and *not* the value that has been set on the microtome. Thickness will have to be measured (see section 11.1 Measuring section thickness).

If the distances between the Disector probes along the x- and y-axes are 200 µm, one Disector probe is placed in an area of 200 × 200 μm^2 = 40000 μm^2. If the Disector probes measures 10 by 10 µm along the x- and y-axes, i.e. 100 μm^2, the probes will include only $100/40000^{th}$ = $1/400^{th}$ of the total area of the section. The area sampling fraction (*asf*) is 1/400, and the inverse of this fraction is 400.

If the series contained only every 10^{th} section, the section sampling fraction (*ssf*) is $1/10^{th}$, and the inverse of this fraction is 10.

If we counted ΣQ^{-} = 200 cells in all the Disector probes, we know that this is ½ × 1/400 ×1/10 = $1/8000^{th}$ of all the cells that there are. If we multiply the 200 cells with the inverse of the sampling fraction, i.e. 8000, we obtain an estimate of total cell number – 1.6 million.

$$200 \times \frac{1}{\frac{10}{20}} \times \frac{1}{\frac{100}{40{,}000}} \times \frac{1}{\frac{1}{10}} = 200 \times 2 \times 400 \times 10 = 200 \times 8{,}000 = 1{,}600{,}000$$

The number or volume of Disector probes that were used to generate the sample actually never enter into this equation. If the objects that are counted can be recognized directly, e.g., serotonin immunoreactive cells of the raphe or hippocampal CA3 pyramidal cells that can be told apart from other cells in the brain, the border of the area that we counted in does not matter. The sampled area could be extended by vast numbers of Disector probes beyond the raphe or beyond CA3 of the hippocampus to encompass the entire country of residence of the investigator. The vast majority of probes would return counts of 0 and not alter the number estimate. It is therefore not necessary to outline the area to be sampled very precisely.

It is different if the objects of interest cannot be recognized directly. A cortical layer III pyramidal cell in V1 at an x100 magnification may be difficult to tell from one in V2. In such cases, an outline needs to be defined that allows the identification of the objects of interest.

The outline may be defined based on regional cytoarchitecture at low magnification or based on other stains in parallel series that allow its definition (e.g., Gritti et al., 2006; Morgan et al., 2014). In these and similar circumstances, the area sampled needs to the defined as good as possible along the parts of the border that have the potential to add to the counts. For example, a cortical area would need to be delimited as good as possible from neighboring areas but not necessarily against the surface/pia mater or white matter. If one were interested in oligodendrocytes, one would, of course, also need a precise outline towards the white matter. Correct outlines may be needed to allow the identification of the objects, but, again, the area of the outline or the numbers of probes placed within it do not enter into any calculation.

N_V × V_{Ref} – Another approach to obtain an estimate of total numbers needs both the volume contained in the Disector probes and the volume of the region of interest. If 200 cells were counted in 100 Disector probes that measure, as in the example above, 10 × 10 × 10 µm or 1000 µm^3, an estimate of number density, N_V, could be 200/(100 × 1000 µm^3) or 2/1000 µm^3. If number density is multiplied by a reference volume, V_{Ref}, i.e. the volume of the same region of interest used in the example above, an estimate of total number is obtained. If the volume of the region of interest is estimated to be, e.g., 0.35 mm^3 or 355,000,000 µm^3 we obtain an estimate of total number of 710,000 cells.

$$710{,}000 = \frac{200}{100 \times 1{,}000\ \mu\text{m}^3} \times 355{,}000{,}000\ \mu\text{m}^3$$

Something does not work here. Using the Fractionator estimate we obtain an estimate of 1.6 million cells, but using the volume of our structure of interest and the $N_V \times V_{Ref}$ approach we obtain only 710,000 cells. The discrepant estimates can be caused by the fact that some Disector probes may not be contained completely within the region of interest but instead fall on the border of the region of interest (this *must* happen sometimes – see below). The parts of the Disector probes that lie outside the border cannot contribute to the counts, while their volume still contributes to the estimate of number density. How this can be accounted for has already been illustrated in Figure 17. While counting, one can keep track of how much of each probe is inside the region of interest. The easiest way to do so is to keep track of the number of corners of the counting frame that fall within the region of

interest. Depending on the degree of precision that is desired, one may choose one, two or more corners to represent the counting frame. In Figure 17, two opposing corners were selected. Although nine counting frames are applied to the region of interest, only eight of the 18 possible selected corners fall into the region of interest. As two corners represent one counting frame the nine frames applied to the region of interest represent only 4 full frames, i.e. only 4/9 of the volume of all the probes applied was contained within the region of interest. If this is corrected for, an estimate of 1,597,500 cells is obtained.

$$1{,}597{,}500 = \frac{200}{\frac{4}{9} \times 100 \times 1{,}000\ \mu m^3} \times 355{,}000{,}000\ \mu m^3$$

Note that within each Disector – regardless of the Disector itself being counted as 1, ½, or 0 – we count all objects that are located within the region of interest and that comply with the counting rules of the Disector probe.

The small difference that remains between Fractionator and $N_V \times V_{Ref}$ estimates (2500 cells) is, in this case, due to a rounding error in the calculation of the reference volume to be used in the example. It may, during a real estimation procedure, also originate from the (lack of) volume estimate precision or the estimation of the number of counting frames that were contained in the region of interest. Across estimates in a number of individuals, the means of Fractionator and $N_V \times V_{Ref}$ estimates should be statistically identical.

As pointed out above, counting frames that fall onto the border of the region of interest cannot be avoided. If we decided to only include Disectors that are completely contained within the region of interest, objects contained within the region do no longer have equal chances to contribute to the count. The problem is illustrated in Figure 22. The object close to the border of the region of interest would only rarely or even never be included in the count, whereas objects that lie farther away from the border have a chance to be counted in many differently positioned counting frames. We can only provide the objects with equal probabilities to be counted if we permit counting frames that are not completely contained within the region of interest. This is a requirement for *both* the Optical Fractionator and the $N_V \times V_{Ref}$ method.

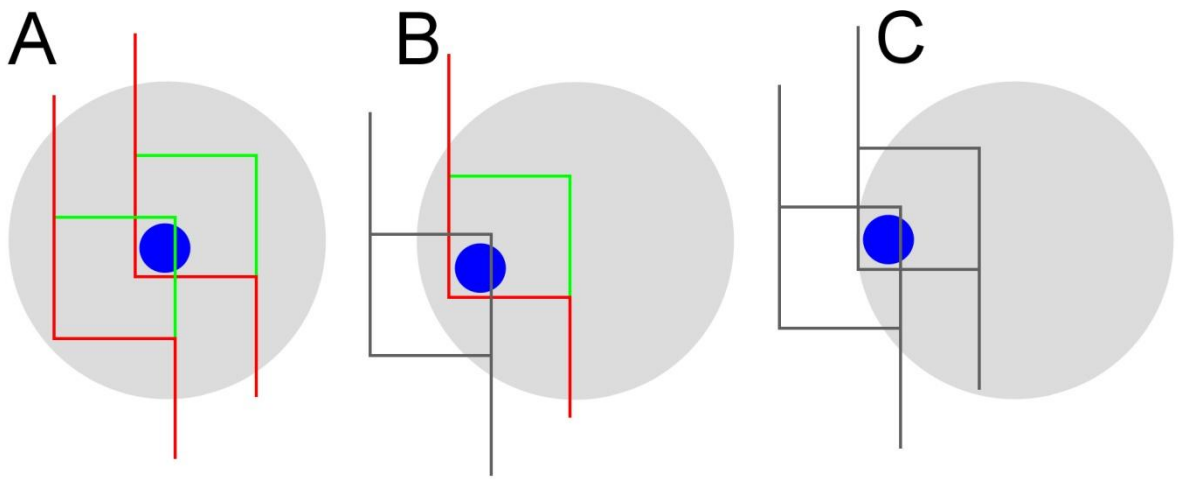

Figure 22 Counting frames that are not fully contained in the region of interest must be used. An object located fairly centrally in the region of interest (**A**) can be contained in many differently positioned counting frames (only two examples are drawn). If counting frames that are not fully located in the region of interest are not used (grayed out frames), the number of frames that could contain a more peripheral object decreases (**B**) until the object cannot be contained in *any* frame placed on the region (**C**). Number would consequently be underestimated.

There is one more thing that needs to be paid attention to when the $N_V \times V_{Ref}$ approach is used – a form of the reference trap. Errors may sneak into the calculation of the reference volume and generate errors in the estimate of total number. For example, sections may be cut at 40 µm thickness but shrink to 20 µm after mounting, drying and coverslipping. During shrinkage, the cells come closer together, and the N_V estimate may double. The reference volume needs to be calculated based on the section thickness at the time of counting, i.e. based on 20 µm and *not* 40 µm. There may be several steps between reference volume estimation, maybe by way of *in vivo* MRI imaging, and cell counting after maybe dehydration, embedding, cutting, rehydration, staining, dehydration, clearing and mounting. Dimensional changes that can affect N_V or V_{Ref} estimates need to be monitored throughout (see section 11.6 *Shrinkage*) and adjusted for.

8.6 Simple Disector implementation

Possibilities to move along the x- and y- axis have already been described in section 7.6 *Simple Cavalieri estimator implementation*. For z-axis measurements of section thickness or Disector height, a microcator can be attached to the microscope. This device will return measurements of z-axis movements of the microscope stage with small µm-fraction

precision. Microscopes, in which the z-axis movement is motorized, may offer a read-out of the z-axis movement that is sufficiently precise to omit a microcator. Precision should be tested, in particular if the microscope has some age, which may cause wear-and-tear on the z-axis drive and possibly deviations between the displayed and the actual movements of the stage. Lastly, even the z-drive knobs of fully manual microscopes may be calibrated to permit z-axis movements of a defined size (Korkmaz and Tümkaya, 1997; Xavier-Vidal, 2010).

If a simple camera setup is available, ImageJ extensions that generate counting frames are available to analyze image stacks. To count live, a calibrated counting frame can be drawn onto a transparency and taped to a screen. Microscope eyepiece reticules that show grids useable as counting frames are available for live counting, although one should be aware that accommodation of the eye, i.e. our ability to focus at objects at different distances from the eye, may extend Disector height beyond the z-distance traveled by the microscope stage.

8.7 A comment on Abercrombie's methods

As pointed out in section 4 *The Problem*, a profile count obtained from sections could be converted into an estimate of cell number using Abercrombie's method (1946), if it was known how likely the cells are to be cut during sectioning. The likelihood to be cut depends on the height of the cells in relation to the thickness of the section. It is therefore necessary to measure cell height, which is not an easy task. Instead, Abercrombie suggested the diameter of the cells to be used as a substitute for cell height. Abercrombie was well aware of possible sources of error, but accepted them for the sake of feasibility and because the error seemed much smaller than the errors possibly associated with earlier methods. Height measurements of a number-weighed, i.e. Disector-sampled, selection of cells would allow such measurements but, as discussed in detail by Hedreen (1998b), this would require all the steps that are needed to generate a Disector-based number estimate and render the measurement of cell height for an Abercrombie estimate redundant.

Prior to the introduction of the Disector, Abercrombie's method was the method of choice in many studies that took quantitative morphology seriously. Sadly, the issue of Abercrombie versus Disector dominated many early methodological discussions. The championship fight between incumbent and contending method unfortunately diverted energy and attention from addressing the fact that many approaches that pretend to generate information about

object number are far, far worse than Abercrombie's method. While the champions were discussing year after year, the spectators were running off with the pretty numbers.

Interestingly, Abercrombie also describes a method to estimate cell numbers that is based on counts in sections of different thicknesses (Abercrombie, 1946). If the count obtained in a thin section is subtracted from the count obtain in a thick section, the difference between the counts would be an unbiased estimate of cell number in a section of a thickness that correspond to the difference in thickness between the sections. The Disector is a special case of Abercrombie's second method, in which the thin section would have a thickness of zero. It corresponds to subtracting the count in the top plane of an optical Disector, from a count of all objects, including those in the top plane, in the Disector sample. This is the same as excluding objects in the top plane from the count. Although the method has been re-introduced several times (Collan, 1998; Ebbeson and Tang, 1965), I am not aware of application in the neurosciences.

9 Length and surface estimators

9.1 Orientation sensitivity of probe-feature interactions

Critical for the understanding of length and surface estimators is the orientation sensitivity of probe-feature interactions when length and surface are estimated by area and line probes (Figure 23). If we probe for the length of, e.g., axons (blue in Figure 23A) running more or less parallel in a fiber tract, a probe area (green in Figure 23A) that itself is parallel to the axons will never be intersected by the axons, i.e. there are no probe-feature interactions that can be counted. The number of area-axon intersections will increase as the angle between the probe and the axon increases (Figure 23B) until we obtain the maximum amount of probe-feature interactions/area when the probe area is perpendicular to the axons (Figure 23C). The number of interactions per area unit is dependent on orientation *and* the length of the axons instead of just the length of the axons. It is only if all orientations of the test area would be equally likely to occur that the number of interaction between the area and axons would be directly proportional to the length of the axons. The roles of probe and feature can also be exchanged in Figure 23. We could use test lines (blue) to probe for, e.g., cortical surface (green), and we face the same problem of orientation-dependence of line-surface interactions.

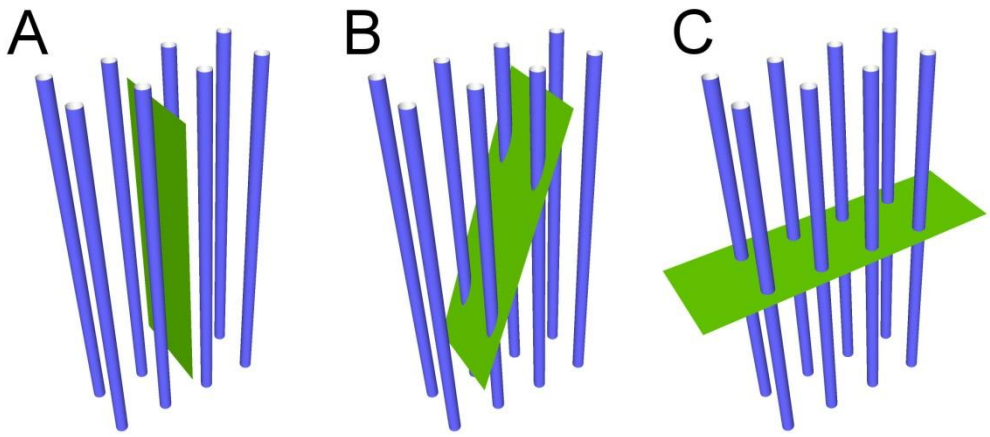

Figure 23 Orientation effects on interactions between a length and an area. The number of interactions between area and length is sensitive to orientation. Even though the blue length and green area are identical in all three images, 0 interactions are observed in (**A**), 4 in (**B**) and 8 in (**C**). This orientation sensitivity applies to estimates of length using area probes and to estimates of area using line probes.

The situation in Figure 23 is of course an extreme case. Locally, the orientations of neuronal processes may appear more or less random, but they are, of course, not (Figure 24A). If the directions of axons were truly random, they would have to, on average, end exactly where they start. This would defy their purpose – connecting a neuron at point A with a neuron at point B. If orientations are completely random they are also called isotropic orientations. The very names that are applied to many structures, e.g., cortical columns and barrels, or bi-tufted (Figure 24B), chandelier or pyramidal cells imply some sort of spatial organization of the cells and their processes. Rather than the randomness of isotropy, we should expect some orderliness, i.e. anisotropy, in the brain.

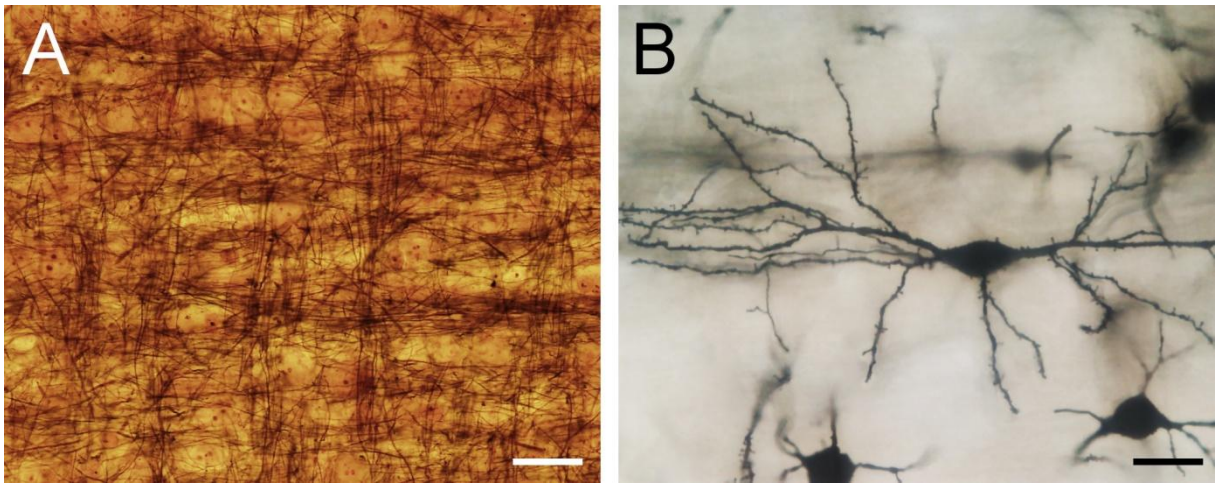

Figure 24 Isotropy and the structure of the brain. Randomness and the structure of the brain do not go well together. (**A**) Axons in the deep layers of the somatosensory cortex of the bank vole (*Myodes glareolus*) in bundles that travel roughly perpendicular to each other between the neurons. Nauta stain. (**B**) Oriented dendrites of a sparsely spinous bi-tufted interneuron (perhaps a Martinotti interneuron) of the Sprague-Dawley rat somatosensory cortex. Golgi stain, montage of several focal planes. Scale bars: 25 µm

Probing the nervous tissue for length and surface requires isotropy, yet we have to expect anisotropy. The problem can be solved by randomizing either the orientation of the structures of interest during tissue preparation or by randomizing the orientation of the probe. Randomizing the orientation of the structures in the sections would require that the orientation in which the tissue is cut is randomized. At a first glance, this is really not a very attractive prospect. However, it is very easy to count probe-feature interactions in randomly oriented sections. Approaches to estimate length and area in sections of (partly) random orientation are therefore described in section 9.4 *Length and Surface estimates in isotropic*

or vertical sections. Randomizing the orientation of probes within sections has become the preferred approach, because it retains the free choice of the orientation of the section. Thick sections are required because sufficient space is needed within the sections to allow the placement of line or area probes in any orientation, i.e. also along the z-axis, which typically limits the size of the probes that can be applied.

Two slightly different approaches have been developed to place randomly oriented area probes within sections. The use of *Isotropic Virtual Planes* (Larsen et al., 1998) refers to changing orientations of probe planes at different sampling locations. This method is computationally demanding and the counting rules are not simple. Subsequently, a method was developed that places a spherical area probe at each sampling location (Mouton et al., 2002). The probe has become known as a Spaceball and is, together with its application, described in the following section.

9.2 Using Spaceballs to probe for length

A spherical probe or spaceball fulfills the requirement of equal chances for all possible orientations, or isotropy, because the surface of a sphere presents all possible orientations of a surface in space. A representation of the surface of a spaceball can be generated by a computer interfaced to a microscope as the focus is moved through the depth of the section at each sampling site (Figure 25). Probe feature interactions occur each time the structure of interest touch or cross the circles that represent the surface of the spaceball.

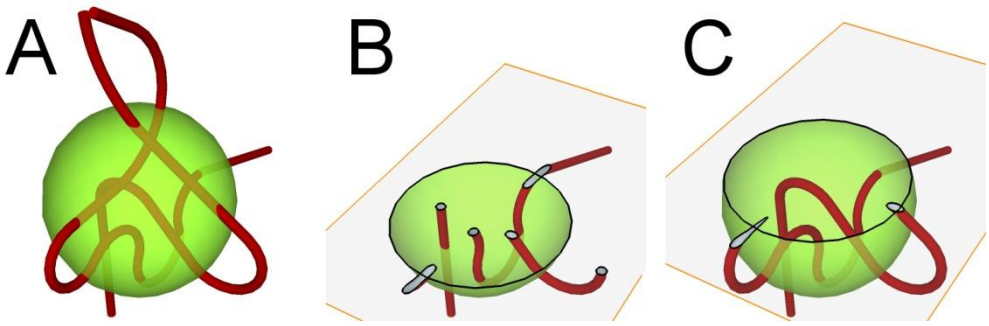

Figure 25 Interactions of a length with a spherical probe. (**A**) A spaceball, i.e. a spherical probe, interacting with two tubular structures. (**B**) and (**C**) represent two views of the sectioned spaceball, corresponding to the movement of the focal plane (grey) in a thick histological section. The tubular structures intersect the surface of the spaceball twice in each of the two section planes. The black outlines of the spaceball surfaces in (**B**) and (**C**) would appear as circles as the focus plane is moved through the depth of a histological section (see Figure 26).

Each orientation in space is not only represented once, but twice on the surface of a spaceball – once each at opposing sides of the spaceball. That means that the application of a hemisphere is sufficient to probe the tissue. In practice, this has the nice effect that a larger probe area can be used at each sampling location. If 20 µm thick sections are probed, the largest sphere that can fit inside the section has a radius of 10 µm and a surface area of 1257 µm^2 ($A_{sphere} = 4\pi r^2$). Instead, we can place a hemisphere with a radius of 20 µm and a surface area of 2513 µm^2 at the same sampling site. In principle, a hemisphere could even be partitioned into multiple slices at each sampling site (Mouton et al., 2002). While this would allow for a further increase in the available probe area, it would also complicate the counting (see below).

A hemisphere was used as a spaceball in Figure 26 to count intersections of tyrosine hydroxylase (TH) immunoreactive axons with the spaceball surface. Small circles appear close to the top of the hemisphere that, at first, quickly increase in size as the focus is moved downwards (Figure 26, 0 to 3 µm). Close to the equator of the hemisphere, the size of the circles increases very little (Figure 26, 8 to 10 µm). In contrast to the counting frame, touch and cross are not equivalent interactions but may be differently influenced by the thickness of the structure of interest (see below). In Figure 26, to cross the surface, i.e. to cross the circle in the 2D representation that we see, was used as a counting criterion. The rapidly changing diameter of the circles close to the top of the spaceball may make it difficult to observe the actual crossing of the structure of interest with the circle. If it was seen once on the outside and once on the inside, it must have crossed in the meantime and an intersection should be counted (e.g., Figure 26 at 1 and 2µm).

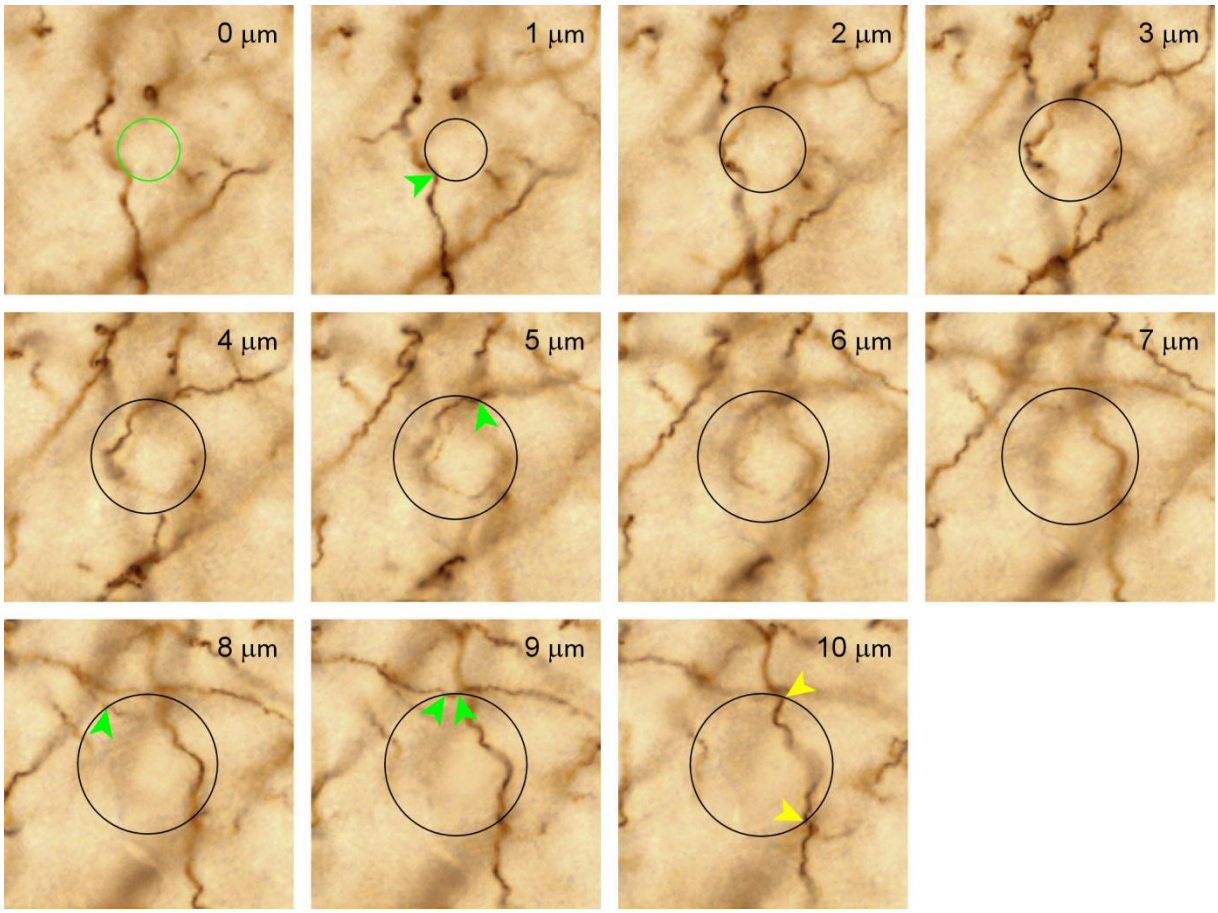

Figure 26 Estimating length using spaceballs as probes. A hemispherical spaceball (black circles) with a radius 10 µm inside a 10 µm deep image stack of neocortex containing tyrosine hydroxylase (TH) immunoreactive axons. Intersections of the axons with the surface of the space ball are marked with arrows. At 0 µm the spaceball only touches the image plain and is not yet visible as a cross-section. The green circle was added to help identify axons that are in focus at 0 µm and, at 1 µm, inside the spaceball. Such axons must have intersected the surface of the spaceball and should be counted (no such cases are observed). At 9 µm, an axon briefly enters the spaceball to immediately leave it again. It must therefore intersect the surface twice (see object C in Figure 27).The two intersections observed at 10 µm would also be visible in the very first plane of the hemisphere completing the spaceball to a sphere. The intersections marked with yellow arrows are therefore only counted as 0.5. A total of 6 axon-spaceball intersections (5 green, 2 yellow) are observed.

Note also that an intersection observed in the very last focal plane that contains the spaceball would also be counted in the very first focal plane of the hemisphere that continues the spaceball. Intersections in the last focal plane should therefore only be counted as 0.5 to avoid over-counting. Alternatively, one could declare one half of the last circle and "exclusion semicircle". If the spaceball was sliced further to increase the available probe area, any probe-feature interactions occurring at the edges of the slices would also

have to be counted as 0.5. Using the crossings of the structure of interest with the surface of the spaceball is rather straightforward if the structures are thin. The decision if an intersection is present or not may not always be unequivocal if the structures get thicker (Figure 27). In this case it is helpful to use the imagined center of the structure of interest to judge if one or more intersections took place (Mouton et al., 2002).

In Figure 26, we count six interactions of axons with a hemisphere that has a radius of 10 µm and a surface area of ($4 \times \pi \times 10^2$) / 2 = 628 µm^2. Following the relationship equation for length using an area probe, we obtain a length density estimate, L_V.

$$L_V = \frac{2Q}{A} = \frac{2\sum Q}{N_{samples} \times area\ per\ sample} = \frac{2 \times 6}{1 \times 628\ \mu m^2}$$

in which Q is the number of intersections counted with the spaceballs, and A is the area that was used to probe the tissue. Usually, there should be more than six interactions, and, of course, we should probe our region of interest with more than a single spaceball. If we had counted 84 intersections using this probe at 53 sampling sites, L_v would be (2 × 84) / (53 × 628 µm^2).

The length density estimate can be converted to the total length by multiplying L_V with the volume of the region that contained the structures of interest.

$$L = L_V \times V_{ref}.$$

Note that of the µm^2 in the denominator of L_V and the µm^3 as the unit of V_{ref} only a unit of length, µm, remains. Note also that, similar to the calculation of the $N_V \times V_{ref}$ estimate of number, we need to keep track of how many of the spaceballs actually were located inside the reference volume. This is tricky if we would have to use the spaceball itself to find out, but one or more arbitrarily placed points can be included with the spaceball probe – often the corners of a square drawn around the spaceball. The number of points would tell us how many spaceballs were contained within the reference volume.

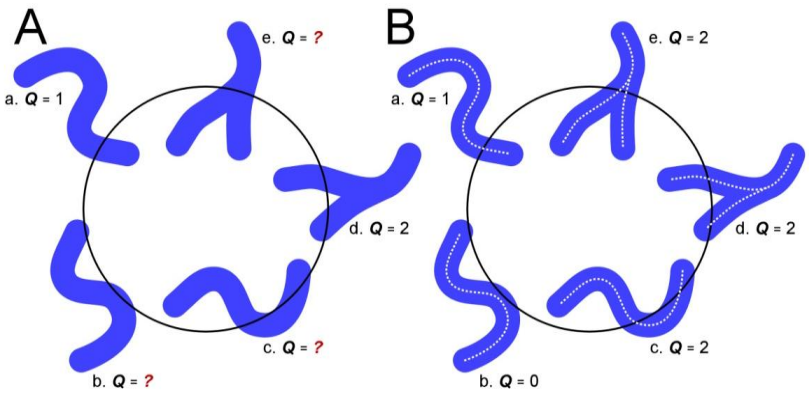

Figure 27 Sometimes it is easy to decide how many interactions need to be counted based on the appearance of the objects of interest alone (cases a and d in A). Other cases are more doubtful (b, c and e in A). If one does not want to count touches as interactions between the spaceball and the length that is to be estimated, imagining the splines (white broken lines in B) may be helpful. Even though the object in case c touches the circle only once, the spline of this object crosses the spaceball twice. We need to count two interactions. Instead, the object in case b touches the spaceball twice, but its' spline does not cross the surface of the spaceball. No interactions are counted in this case. Even though only one structure crosses the spaceball in case e, the spline crosses twice and two interactions need to be counted.

Alternatively, a fractionator approach can be used to calculate total length. Getting at the formula, which is simple enough, is unfortunately a little more convoluted than was the case for a Fractionator cell count. When L_V was multiplied with V_{ref}, we nicely got rid of the area in the denominator of L_V and conveniently ended up with a length. Multiplying by the dimension-less inverse of the sampling fraction does not help us to get rid of the area or to end up with a length. L_V therefore needs to be converted to a length before we can multiply it with the inverse of the sampling fraction. The conversion is stunningly simple. We can convert L_V to a length by multiplying L_V with the volume of *arbitrarily* sized boxes that we place around the spaceballs. The most conveniently sized box is the volume of the tissue that is probed with each spaceball, V_{probed}. In the plane of the section this box has an area of A_{step}. The depth of the box would be section thickness, h. The number of boxes that we looked at now corresponds to the number of sampling sites, $N_{samples}$.

$$L\ in\ probed\ volume = N_{samples} \times V_{probed} \times L_v$$

The dimensions of this box are also used to calculate the area and thickness sampling fractions. The chosen size of the box is convenient because its area equals A_{step} and its

height equals h. Area and thickness sampling fractions are 1 and can be ignored. Only the section sampling fraction, ssf, needs to be known. Length can now be estimated from parameters that we know and the counts.

$$L = \frac{1}{ssf} \times N_{samples} \times V_{probed} \times L_v$$

$$= \frac{1}{ssf} \times N_{samples} \times V_{probed} \times \frac{2 \sum Q}{N_{samples} \times area\ per\ sample}$$

$$= \frac{1}{ssf} \times V_{probed} \times \frac{2 \sum Q}{area\ per\ sample}$$

$$= \frac{1}{ssf} \times A_{step} \times h \times \frac{2 \sum Q}{area\ per\ sample}$$

Note that, the number of samples or the total volume probed do no longer appear in the final equation. As long as probe feature interactions are only counted within the region of interest, it does not matter if much of the volume associated with each sampling location would be outside the region of interest when the probe itself is inside the region and vice versa. Note also that the most complicated mathematical operations required are simple divisions and multiplications, and not even many of them. The square roots are yet to come.

Figure 27 showed that it sometimes may be difficult to decide if a structure intersects the probe area or not. This is because biological structures are not infinitely thin mathematical lines but have a thickness – invisibly thin would really not make things easier. That there is a thickness of an axon or a vessel may result in a bias if we use “touching the probe area” instead of the preferred “intersecting the probe area” as the counting criterion. E.g., structure b in Figure 27B touches twice but does not intersect. This may bias estimates of length, and the size of the bias depends on the thickness of the structures in relation to the diameter of the spaceball. The size of the bias can be defined for simple shapes in the form of a factor $1 + 0.25(d^2/t^2)$ (factor for hemispheres; Mouton et al., 2002), in which d is the diameter of the structure of interest and t the radius of the space ball. If, e.g., the length of 2 µm thick axons or 6 µm thick capillaries are estimated with spaceballs of a radius of 20 µm, the factors would be 1.01 and 1.09. We see 1.01 times (1%) more touches than there are actual intersections of the axon with the area of the spaceball and 1.09 times (9%) more touches of capillaries than there are actual intersections. Even though the bias is not large, it may be avoided by using “intersect” instead of “touch” as counting criterion.

9.3 Probing Surface: Isotropic Line Estimators

Conceptually the roles of probe and feature of interest introduced for length estimators are exchanged when surfaces are probed, while the problem of orientation sensitivity of the probe remains (Figure 23). The solution is, again, a randomization of the orientation of line probes if we want to remain free to choose the orientation of the tissue sections that we like best. Two methods using a randomization of line orientations have been presented. Similar to *Isotropic Virtual Planes* (Larsen et al., 1998), the *Isotropic Fakir* (Kubínová and Janácek, 1998) uses randomly oriented sets of straight lines. Similar to the spaceballs (Mouton et al., 2002), *Virtual Cycloids* (Gokhale et al., 2004) use test lines that each incorporate multiple orientations. Both methods are applied as virtual probes to thick tissue sections.

The easiest way to generate a line with a random orientation is to decide on a point of origin and, thereafter, randomly select two angles that would correspond to the longitude and latitude used to specify positions on the surface of the earth (Figure 28A). Like all other relationship equations, surface density estimates require that the number of interactions between the line probes and a surface is directly proportional to the length of the lines and surface area and nothing else. Unfortunately, the simple selection of two angles does not yet fulfill this requirement. In Figure 28A, areas that are associated with a degree of longitude and a degree of latitude are much narrower close to the poles than they are close to the equator. Because of this, countries far up north or down south look – bolstering national self-esteem – much larger in some maps than they in reality are. Also, the relatively small area close to the poles would interact more often with randomly oriented line probes than the larger areas close to the equator (Figure 28B), and we would erroneously confirm that some countries are much larger than they are. This can be compensated for by sine-weighting the angle for the latitude of the probe lines. Instead of picking a random angle, a random sine value (between zero and one) is selected, and the corresponding angle is calculated and used (Figure 28C). Using sine-weighted angles, line probe-surface interactions distribute evenly over the globe.

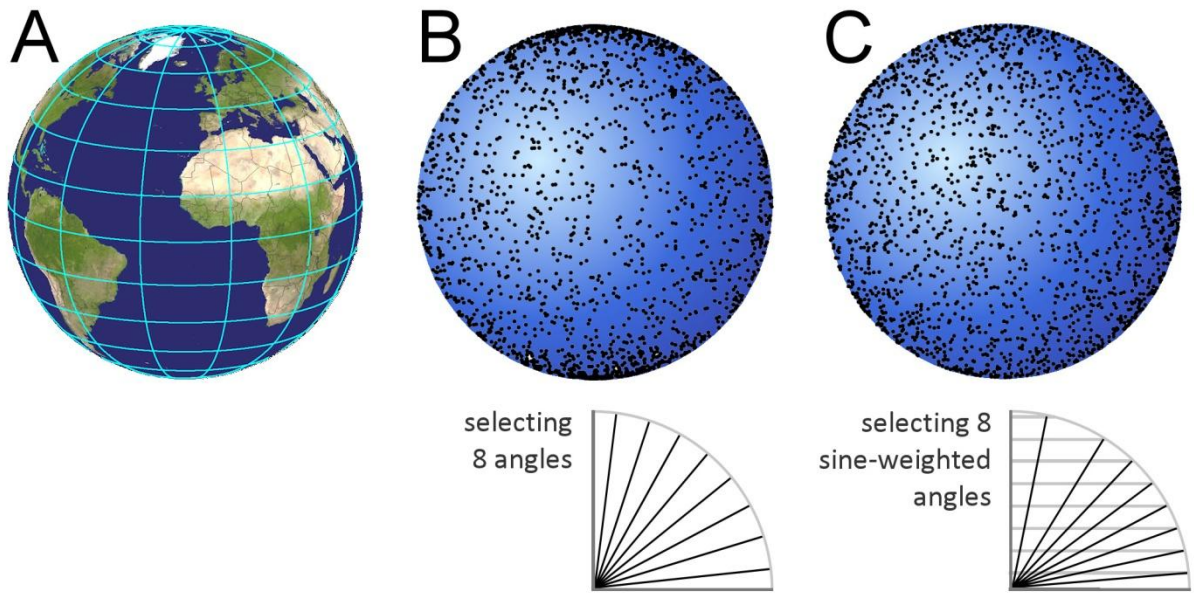

Figure 28 Interactions of randomly oriented lines with a surface (**A**) Using the common latitude and longitude coordinate system, the areas associated with a degree of latitude and a degree of longitude are much smaller close to the poles than close to the equator. (**B**) Random latitude and longitudes are selected for lines originating from the center of a sphere. They interact more frequently with the surface of the sphere close to the poles than with the surface close to the equator. (**C**) Randomly selected longitudes are combined with randomly selected sine-weighed latitudes. The interactions between the lines and the surface of the sphere are now evenly distributed over the surface. Angles for latitudes below (**B**) and (**C**) show eight uniform random systematic angles in (**B**) and eight uniform random systematic sine-weighed angles in (**C**). (Globe in (**A**) generated using satglobe4 (Kleder, 2005))

A sphere and an origin of the lines from the center of the sphere were chosen to illustrate the problem and its solution. They are also valid when the shape is not a sphere and when the origin of the lines lies outside the structure of interest.

While sine-weighting does allow the placement of randomly oriented lines, it still requires the placement of many differently oriented lines at the sampling sites. In the same way in which a spaceball, used to estimate length, contained all orientations of a surface in space, some convenience can be gained by using a line which contains all orientations in a sine weighted manner. Such a line is called a cycloid. A cycloid would be formed by a point on a rolling circle (Figure 29). A cycloid's steep initial segment is rather short (corresponding to fewer lines going to the polar region) compared to the longer, shallow-angled approach to its' peak value (corresponding to more lines going to the equatorial region).

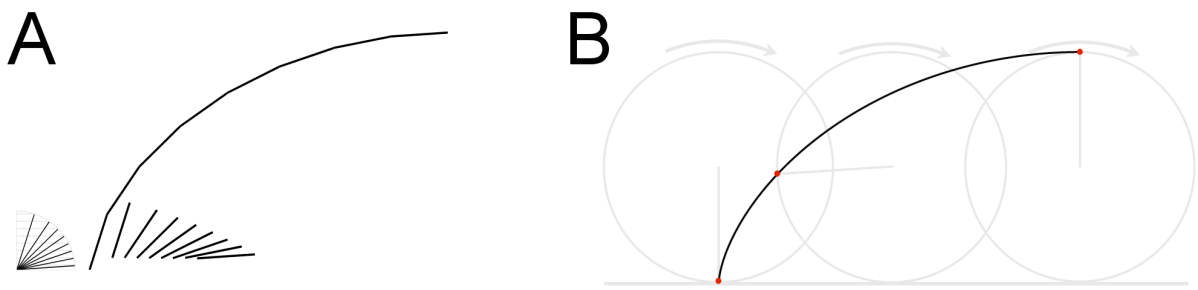

Figure 29 Cycloids. (**A**) The curved shape that if formed by attaching eight sine-weighted lines (Figure 28C) of equal length to each other resembles a cycloid. If increasingly more and shorter lines were used, the shape would be indistinguishable from the cycloid in (**B**), which is the curve formed by a point on a rolling circle.

As a virtual probe, the cycloid will be represented by a point that moves as the focal plane is moved up or down through the section (Gokhale et al., 2004). Note that at different sampling sites, the point will move in different directions because randomly selected longitudes, i.e. random directions in the plane of the section, will be used at each sampling site. A probe-feature interaction is counted each time the point moves across/intersects a surface (Figure 30).

The number of interactions between cycloids pointing to different longitudes and a surface can now be used in the relationship equation for surface density, S_V.

$$S_V = 2I_L = \frac{2I}{length\ per\ sample}$$

in which I is the number of intersections counted. For the one sampling site using two cycloids illustrated in Figure 30, S_V is 10/28µm. The calculation of total surface is analogous to the calculation of total length. We can obtain an estimate by multiplying with a reference volume, $S = S_V \times V_{ref}$. E.g. the volume represented by the stack in Figure 30 is 50 µm × 50 µm × 7 µm, i.e. 17500 µm^3. An estimate of the nuclear surface contained in this volume is (10/28µm) × 17500 µm^3 = 6250 µm^2.

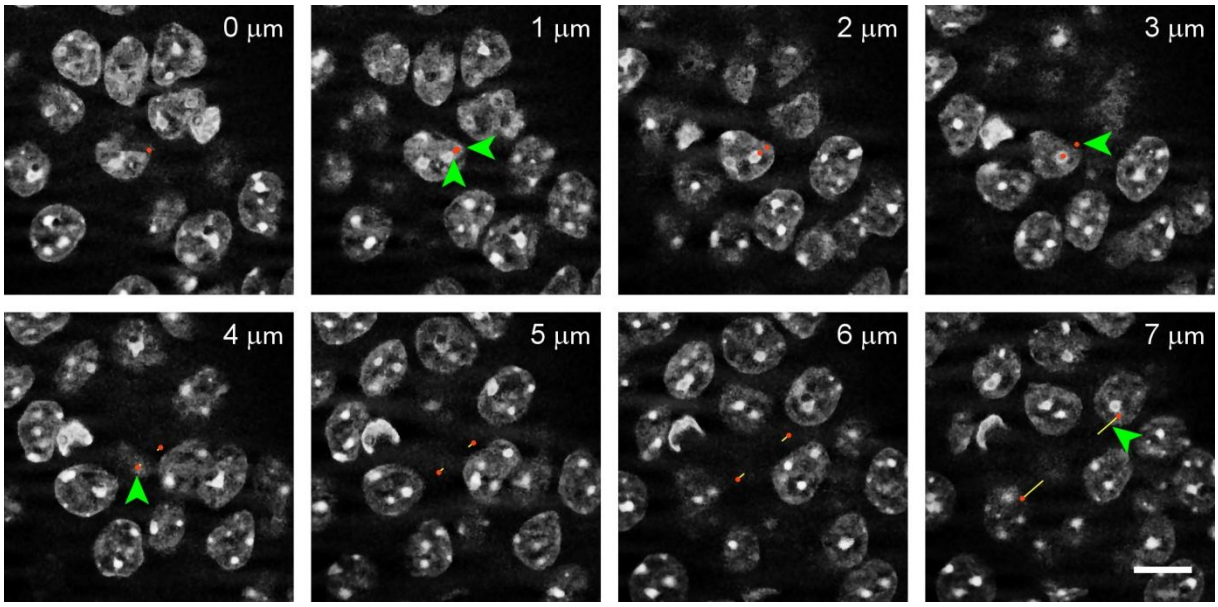

Figure 30 Estimating surface using virtual cycloids. Image stack acquired using structured illumination of mouse DAPI stained hippocampal pyramidal cell nuclei. In each image, the intersections between the image plane and the two 7 µm high virtual cycloids (passing from the center of the image to the lower left and upper right) are marked by red dots. The parameter that one may be interested in is nuclear surface area. The dots pass five times across the nuclear membranes. Both cycloid pass from outside of a nucleus to the inside a nucleus between 0 and 1 µm generating one probe-feature interaction for each cycloid; they are leaving the nucleus again at 3 µm and 4 µm again generating one probe feature interaction for each cycloid. The last interaction is generated as one of the cycloid passes inside a nucleus at 7 µm. The length of a cycloid is twice its height, i.e. 14 µm. Two cycloids are used and the total probe length is 28 µm. Scalebar: 10 µm.

For a Fractionator estimate, S_V needs to be first converted to a surface in the same way in which it was done for length in section 9.2 *Using Spaceballs to probe for length*. In addition to the actual counts, only the sampling parameters and section thickness need to be known to calculate an estimate of total surface.

$$S = \frac{1}{ssf} \times A_{step} \times h \times \frac{2 \sum I}{length\ per\ sample}$$

A problem of virtual isotropic line estimators is the recognition of surfaces in thick tissue sections, which can be difficult if the surface is tangential to the plane of the section. While the surface of a spaceballs is represented by a quickly moving but otherwise distinct line close to the top of the spaceball, this is most likely not the case for, e.g., neuronal membranes or vascular surfaces. If recognizing probe-feature interactions is judged to be too uncertain, surface estimators that are described in following section could be used.

9.4 Length and Surface estimates in isotropic or vertical sections

When isotropic surface or line estimators are used, the orientation of the section in which they are applied does not matter. The probes guarantee that interactions between probe and feature are independent of orientation. The freedom to choose the plane of the section comes at a price. The sections need to be relatively thick, shrinkage along the z-axis ought to be minimized (see section 11.6 *Shrinkage*), and it may, occasionally, be difficult to judge if the probe and feature interact. Although simple and inexpensive ways to implement probes can be devised, their implementation is not as straight forward as one might wish. Some of these difficulties can be overcome by giving up the freedom to cut the tissue in a desired direction. Before the advent of virtual probes in thick sections, this was part of earlier approaches that guaranteed isotropic probe-feature interactions.

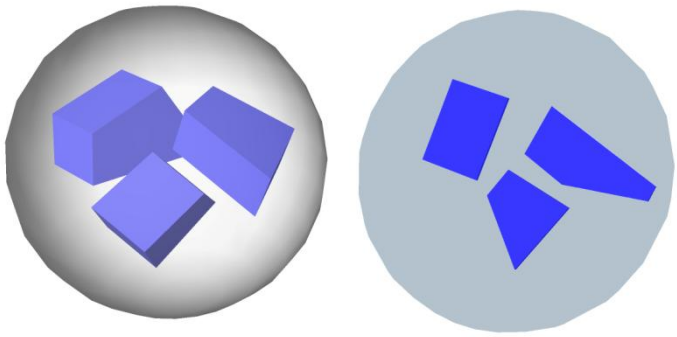

Figure 31 Isotropic section prepared using the Isector. In the form the Isector was originally proposed, the tissue was embedded in a spherical mold. The sphere containing the tissue (left) was thereafter randomly rotated, rembedded and cut to generate an isotropic section through the tissue blocks (right).

The Isector – The most radical approach to guarantee isotropic probe-feature interactions is to prepare isotropic sections, i.e. sections in which the direction of cutting has been completely randomized. The approach has be termed the Isector (Nyengaard and Gundersen, 1992). Using the Isector, samples of the tissue are embedded in random orientations and cut (Figure 31). The sections that are generated are also called *isotropic, uniform random (IUR) sections*. Using the Isector may appear almost ludicrous in an organ in which the orientation of the section may even determine if the region of interest can be recognized or not. However, it is less so than it appears at a first glance. A practical approach for the preparation of Isector samples was outlined by Løkkegaard et al. (2001) in a study that estimated capillary length in the subdivision of the human hippocampus. In this study,

sections of varying thicknesses were cut from larger blocks and used for different stains. Samples for the Isector were prepared by micro-dissecting samples from thick sections that were stained free-floating prior to embedding to allow the identification of regions. The Isector does *not* demand that the entire structure is cut at random. Sections can be cut and stained in the usual manner provided that samples for subsequent embedding/re-cutting can be obtained from these or adjacent sections.

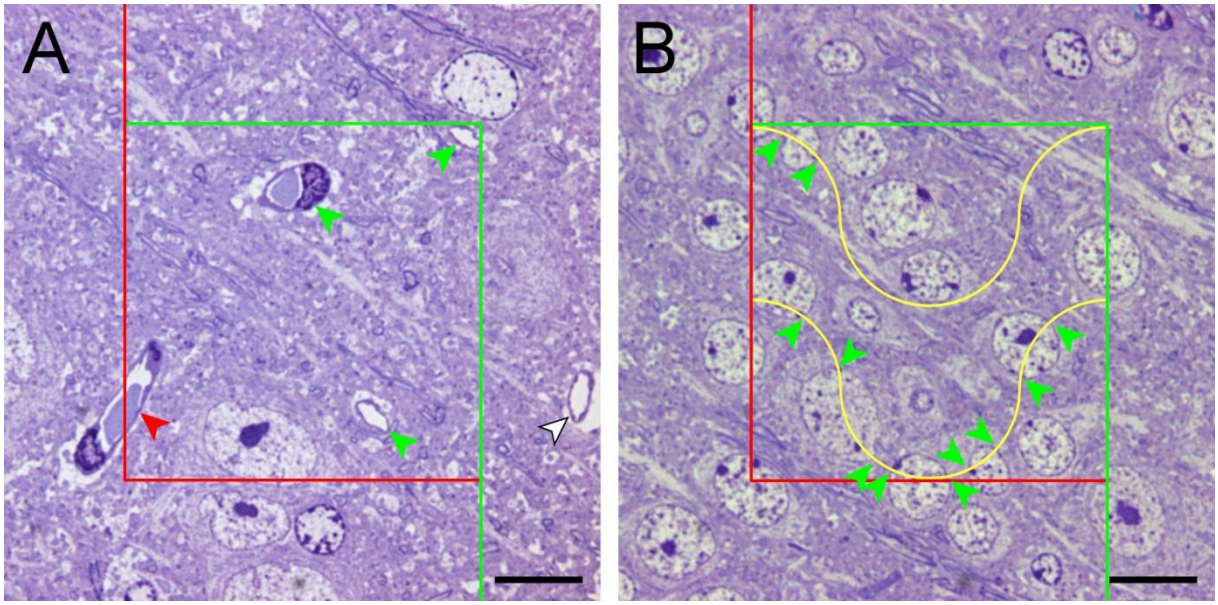

Figure 32 Probing for length and surface in isotropic sections. 0.2 µm semithin sections of Epon-embedded, toluidine blue stained mouse neocortex. (**A**) A counting frame used to define a sample of the area and to count intersections of capillaries with the plane of the section in this area. The three capillary profiles labeled with green arrows can be counted. The profile marked with the red arrow crosses the exclusion line of the counting frame and is not counted. An additional profile is labeled with an open arrow. (**B**) A Merz line grid to count intersections of the lines with the surface of cortical nuclei. The area to which the lines are applied is again defined by a counting frame. Straight lines could also be used, but the hemicircles composing the Merz lines may be more efficient if the surfaces to be estimate have a preferred orientation in the section. Eleven intersections of the lines and the boundaries of nuclei are counted. Scale bars: 10 µm

Because the sections are isotropic, estimates of surface can be obtained by applying probe lines of any orientation to the sections to count line-surface intersections. For length estimates, the plane of the section represents the test area – the way it was illustrated in Figure 8. Estimates of length can be obtained by counting intersections of structures with the plane of the section. For both length and surface estimates, the estimation procedure can be restricted to a sample of the area of the section using the unbiased counting frame

(section *8.2 The unbiased counting frame*; Figure 32). The resulting surface or length density estimates can be converted to total length or surface using either the Fractionator or the N_V × V_{Ref} approach (e.g., Jørgensen et al., 2008; Tang and Nyengaard, 1997).

In the example for length (Figure 32A), three intersections of vessels are counted within 40 × 40 µm = 1600 µm^2 area defined by the counting frame. An estimate of L_V based on just this one probe would therefore be 2 × 3 / 1600 µm^2. The volume of the probe equals the area of the frame × section thickness, i.e. 1600 µm^2 × 0.2 = 320 µm^3. An estimate of the length, L, of vessels in just this one probe would be (2 × 3 / 1600 µm^2) × 320 µm^3 = 1.2 µm.

In the example for surface (Figure 32B), the length of the test lines test lines (two circles with a radius of 10 µm) is ~126 µm. The sizes of the counting frame area and section thickness are identical to those in the example for length. An estimate of the surface, S, of nuclei in just this one probe would be (2 × 11 / 126 µm) × 320 µm^3 = 55 µm^2.

Vertical sections – Vertical sections as a means to guarantee isotropic probe-feature interactions were first introduced for surface estimators by Baddeley et al. (1986) and subsequently also for length estimators (Gokhale, 1990). Vertical sections do not require the complete randomization of the cutting direction. Essentially, half of the randomization is generated during the cutting of the tissue, while the other half relies on randomization during the application of the probes to the tissue. These two randomizations work similar to the randomization of the orientations of test lines already introduced in section 9.3 *Probing Surface: Isotropic Line Estimators*. A random angle for the longitude is chosen to rotate the region of interest or slabs of the region prior to (re-)cutting (Figure 33). Note that the horizontal, i.e. the plane in which we rotate the tissue, can still be chosen freely. The horizontal would correspond to a plane passing through the equator in Figure 28.

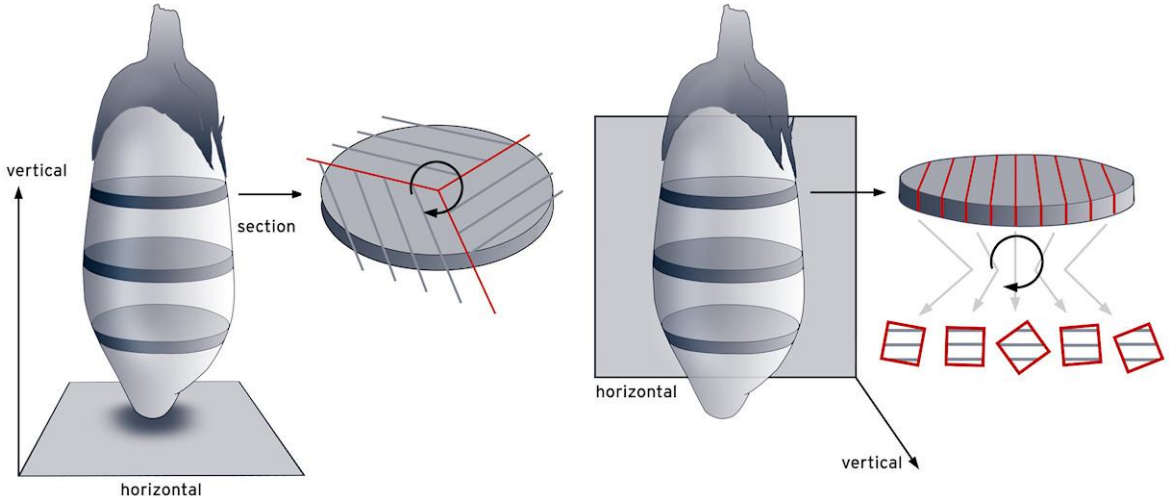

Figure 33 Two-ways to prepare vertical sections from a region of interest. A plane parallel to the equator, the horizontal, can be chosen freely. The choice of the horizontal does not itself impact on the way in which sections are prepared, but only on the way in which the tissue must be rotated to generate the first of the two random angles that are necessary to ensure isotropy in vertical sections.

Next, the rotated tissue is cut into sections perpendicular/vertical (hence, vertical sections) to the horizontal. The sections generated in this way would pass through a randomly selected longitude. They are called vertical uniform random (VUR) sections. The randomization of the angle that would specify the latitude is achieved during the application of the probe lines or probe areas. The major gain, in comparison to sections prepared with the Isector, is that it is possible to at least decide where we place the horizontal, i.e. the plane in which longitude is specified. From an anatomist's point of view, vertical sections are suited to parts of the brain that have a natural horizontal. If we rotate neocortex (as a flat-mount or a cortex block; Dorph-Petersen et al., 2009) in a horizontal that is parallel to its surface and cut it vertical to the surface, we will always see the typical six layered organization of the cortex. The same will be true for any part of the CNS that has a tiered organization in one plane or can be prepared to have one. Other structures may have a natural vertical. Rotating, e.g., the olfactory bulbs around their central axis will provide sections looking fairly similar to sections cut either sagittally or horizontally. Similar to the practical approaches that have been used with the Isector, it is not necessary to sacrifice the entire region of interest to vertical sectioning (Hosseini-Sharifabad and Nyengaard, 2007).

Probe lines applied to vertical sections face the same problem that was already discussed for virtual isotropic lines. Choosing random angles for their application will result in too many lines passing into the direction of the poles. The solution to this problem is, again, a sine-

weighted selection of the angles, i.e. picking a random value between zero and one, calculating which angle would have this value as its sine, and using this angle to apply a probe line to the tissue (Figure 34A). Again, efficiency can be gained by using a line that is a cycloid, i.e. a line representing all orientations in a sine-weighted manner. Intersections between the lines and the surfaces of interest are counted. In contrast to the use of isotropic virtual lines, sections can be thin, because the cycloid now can be placed in the plane of the section instead of perpendicular to it. Probe-feature interactions may be easier to recognize in thin sections. Also, in the critical zone along the long axis of the cycloid, we are looking for probe-feature interactions "from the side" instead of "from the top", which should make it easier to define probe feature interactions when the surface and the probe line are running close to parallel through the tissue. Total surface is calculated using Fractionator or $S_V \times V_{ref}$ approaches.

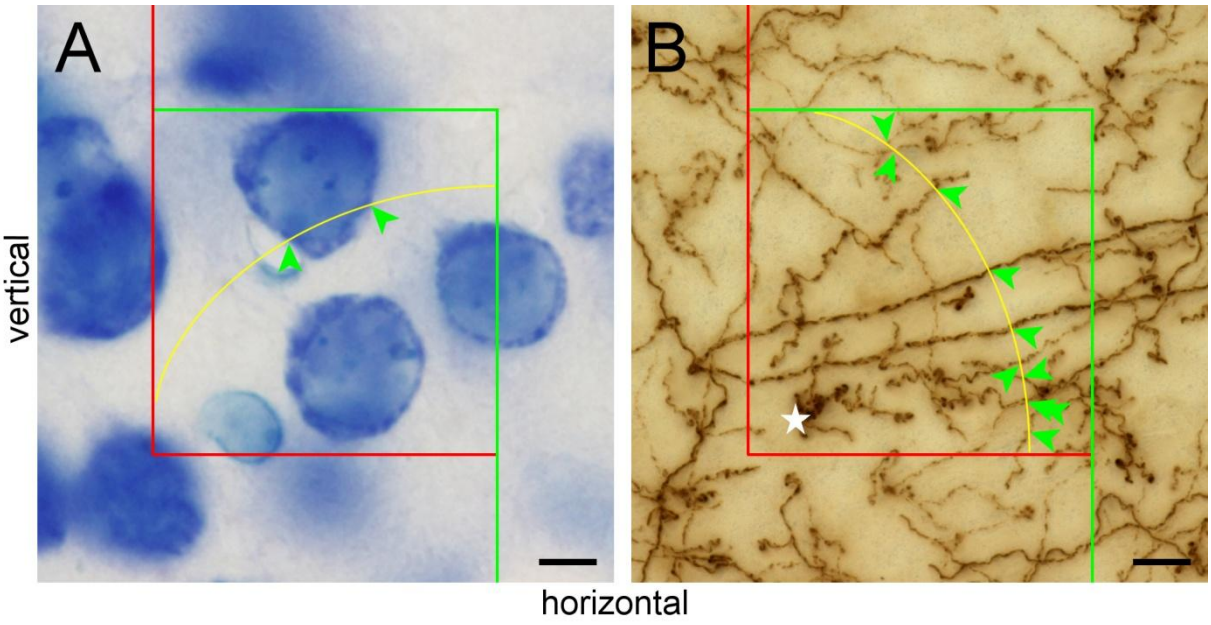

Figure 34 Surface and length estimates in vertical sections. (**A**) A cycloid line probe interacts twice with the surface of a neuron. Scalebar: 5 µm. (**B**) A cycloid area probe interacts 10 times with tyrosine hydroxylase immunoreactive axons in a minimum density projection of an image stack spanning a 40 µm thick section. Image planes were spaced 0.2 µm apart. To resolve interactions between an area probe and axons in dense axon clusters (asterisk) the image stack would have to be examined image by image. Scalebar: 10 µm

In contrast to length estimates using spaceballs, length estimates in vertical sections face the same problem as surface estimates. We therefore again have to apply a weighting to area probes applied to the sections. However, the roles of probe and feature have been

exchanged. Instead of a (length of a) line that will estimate an area, we use an area probe to estimate a length. To compensate for the facts that lengths pointing into the direction of the equator of the vertical section would interact relatively fewer times with a spherical surface probe that lengths pointing towards the pole we need to compensate by providing more area close to the equator. This can be achieved by cosine-weighting the test lines – either by randomly selecting cosine-weighted directions of flat probe areas or by using a probe area that bends into a (surprise!) cycloid shape. While the long-axis needed to be parallel to the horiziontal to estimate surface, this cycloid would have its long axis parallel to the vertical (Gokhale, 1990). The area that we apply would be a projection of the line, whether straight or in the form of a cycloid, through the depth of the section, i.e. the length of the line multiplied by the thickness of the section (Figure 34B). In practice, the focal plane is moved through the depth of the section and intersections of the line (representing the probe area) and structures of interest are counted. Once again, total length can be calculated using Fractionator or $L_V \times V_{ref}$ approaches. Well documented applications of length estimators in vertical sections can be found in, e.g., Artacho-Pérula et al. (1999). Vertical sections are also easily applied to macroscopic *in vivo* imaging data sets, in which the plane in which a structure is visualized can be chosen freely (Acer et al., 2010; Cruz-Orive et al., 2014).

9.5 Simple implementation of virtual length and surface probes

For the selection of sampling sites and the spacing of stack images along the z-axis, some of the simple and inexpensive ways already introduced in the preceding sections may be useful.

If image stacks of sampled locations are available, circles that represent a space ball can be drawn onto the individual images of the stack. If, e.g., 10 µm along the z-axis of the stack contain 21 images (one image at zero µm and 20 images at increments of 0.5 µm) that need to be probed with a spaceball (hemisphere) that has a radius of 10 µm, the sine is divided into 21 equally spaced values separated by increments of 0.05 (0, 0.05, 0.10, 0.15 … 0.95, 1) and the angles corresponding to the values are calculated (0, 2.9, 5.7, 8.6 … 71.8, 90). The radius of the circles that need to be drawn onto individual images will be the cosine of these angles multiplied by the radius of the spaceball (10µm, 9.99 µm, 9.95 µm, 9.89 µm … 3.12 µm, 0 µm). If the graphics software permits the recording of the image manipulation in form of an executable script (even recreational graphics software may allow this), the script can

be applied to the remaining stacks of the sample. Figure 26 was prepared in this manner. It is advisable to keep an unmodified backup of the stacks.

Even isotropic line probes may be implemented without the aid of advanced software. A point could be moved from image to image of a stack by the distances that correspond to the displacement of the point along the x-axis of a cycloid for y-axis increments of the cycloid. The y-axis increments would correspond to the distances between the images in the stack. The x values will be $0.5h \times cos^{-1}\left(1-\frac{y}{0.5h}\right)-\sqrt{h \times y - y^2}$. h is the height of the entire cycloid/image stack and y the height (in the stack) of the image that the x value needs to be calculated for. The length of the cycloid probe is $2h$. Note, that the displacement of the point in the plane of the section should take place in all possible directions. The x-positions for different height in the stack could be marked on a ruler that is applied to the screen as a physical or digital overlay. Working "live" with the microscope one would at least need a precise readout of z-axis movement of the stage.

Length and surface estimators in IUR and VUR sections are back into the realm of very simple implementations. A counting frame that samples the area of an IUR section for intersections of the structures of interest with the plane of the section can be overlaid the image by way of an eyepiece reticules, transparencies taped to a screen, or as a digital overlay of recorded images. Images of cycloids for length and surface estimates in VUR sections are difficult to draw freehand, but suitably scaled ones may be generated with MATLAB and placed over images or taped to flat screens. Also, a cycloid grid plugin was developed for ImageJ.

10 Good enough? – Estimate Precision

A recurrent question pertaining to estimates of number, length, surface or volume is whether they are "good enough". If the tissue has been sampled and probed correctly, statistically valid estimates are obtained. Unfortunately, statistically valid does not necessarily mean the least bit useful. To make numbers useful in quantitative descriptive or experimental work, we ought to be able to put some trust into the numbers. If we report one-hundred thousand cells or thirty meters of capillaries, there should be about that number of cells and that length of capillaries. It is the precision of estimates that determines how good, i.e. useful, these estimates can be. In purely descriptive quantitative studies, the precision indicates how close the reported mean of the estimates is likely to be to the true mean of the sample. In experimental studies, estimate precision in part determines how likely it is that differences between control and experimental groups can be detected statistically. To answer the question if the estimate is "good enough", we need to generate a number that we can use to assess precision in the context of the study. Considering that "not good enough" usually means useless even if the sampling and probing were done correctly in a statistical sense, an understanding of the preceding sections is equally useless if we cannot at least get close to an answer of the "Good enough?" question. Consequently, some space will be spent on estimating and judging precision. There will be quite a few new concepts, and we will finally be challenged by square roots. It may be a good idea to take a breather before setting out.

10.1 What is a *CE*?

A parameter that is useful to describe the precision of an estimate is the *Coefficient of Error* or, short, the *CE*. To understand what the *CE* is, we again need to take a look at why and how an estimate is generated. Typically, estimates are generated because it is not feasible to determine the precise values of the parameters that one is interested in. There are, e.g., simply too many cells in too many sections to count them all. To generate a valid estimate, a statistically representative sample needs to be drawn from each individual, i.e. the mean of the estimates obtained from *all possible* samples will be the true value of the parameter of interest in that individual.

Let us estimate the volume of a brain region. If we would use the Cavalieri Estimator to estimate the area of the brain region in *all* sections that can be cut and, next, plot the point counts obtained in each section along the axis in which the brain was cut, we may obtain a graph like A. This is what we often cannot do because it would require too much work. Instead, we draw a sample of the sections. Because the size of the brain region differs slightly from section to section, each sample will return a slightly different point count. The resulting estimates differ even though we sample the same structure each time.

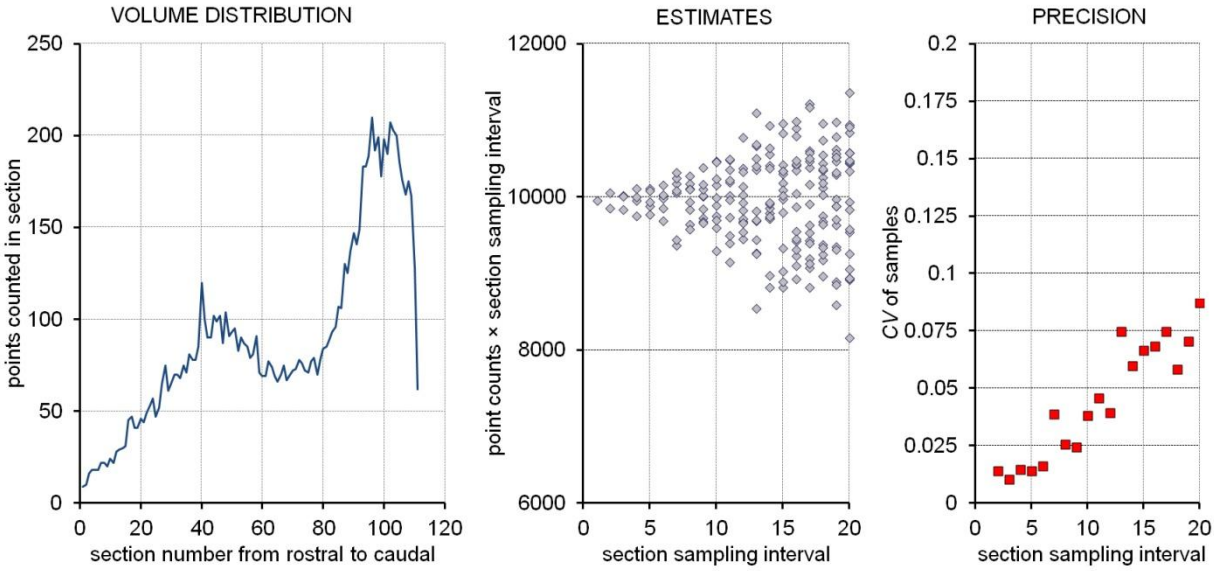

Figure 35 Sampling a volume and the precision of volume estimates. (**A**) Volume distribution of the hippocampal dentate gyrus granule cell layer in an exhaustive series of 20 µm thick coronal sections. (**B**) If subsamples are drawn with increasingly larger distances between the sections, the volume estimates (point counts from the sections scaled by the sampling interval) start to diverge. (**C**) The standard deviation of the estimates generated by a particular sampling divided by the mean, i.e. the *Coefficient of Variation*, is a measure of the precision that can be obtained with the sampling scheme. If, e.g., every 10th section is analysed, the variability of the estimates generated with this sampling scheme amounts to less than 0.05, i.e. less than 5% of the mean.

Figure 35B illustrates the estimates obtained from all samples that used sampling intervals of every 2nd, 3rd, 4th … 12th, … or 20th section. As one may expect, the estimates tend to vary more if fewer and fewer sections are used. If data of this type is available, the *Coeffi cient of Variation* or, short, *CV*, of the samples that belong to a particular sampling interval can be calculated. The *CV* is the standard deviation of all estimates of a particular sampling interval divided by their mean. The *CV* provides a measure of how much the estimates vary relative

to the mean (Gundersen and Jensen, 1987). Figure 35C is a plot of the *CV*s that were obtained from the variability of the estimates in Figure 35B.

CE estimators provide an estimate of the *CV* that is calculated from repeated estimates in the manner described above. Notably, one does *not* need to count in all sections to be able to draw the subsamples that belong to a particular sampling interval. Instead, *several approaches have been developed that allow a CE estimate to be calculated from a single sample*. While the ability to calculate the *CE* from a single sample represents a significant saving in terms of the workload to generate a measure of precision, it does come at a price. *CE* estimators do not provide more than their name promises: an *estimate* of the *CE*. In the same way in which any *one* estimate of, e.g., cell number may deviate from the mean of the group of animals, any *one CE* estimate may not be close to the true precision that is associated with a sampling scheme (Schmitz and Hof, 2000). In neither case is it likely that the final assessment of the data will be based on a single observation, i.e. an *n* of one.

10.2 Why is a *CE* important?

Let us assume that we sample a brain structure that in all individuals of a control group is completely identical in size and shape (black circle in Figure 36). The true value of whatever we are interested in is 100. If the structures are completely alike, this situation would correspond to the repeated sampling of one structure illustrated in Figure 35. Because we sample, we will *not* obtain the same value from each individual but slightly different ones. The standard deviation of the group mean will correspond to the standard deviation of the estimates for the sampling interval that was used. Instead of seeing identical values for all individuals in the control group, they are now represented by the green circles in Figure 36. Let us now add an experimental group, in which the size and shape of the region are, again, identical in each individual. In our experimental group, the region and, therefore, the group mean of our estimates, is slightly larger (black diamond in Figure 36) than that of our control group. In the experimental group, the true value of whatever we are interested in is 110. Once again, because we sample we will see slightly different value for each individual instead of identical ones. The standard deviation of the group mean will again be identical to the standard deviation of the samples obtained with the sampling interval that was used. Instead of identical values, we now obtain the green diamonds in Figure 36 for our experimental group. We were lucky. The difference between the green control sample and

the green experimental sample is highly significant using a two-tailed *t*-test (p = 0.00008). Had we been lazier and sampled less, control and experimental values may look like the red circles and red diamonds in Figure 36. Less precise estimates means more variability in the groups and a smaller chance to detect a group difference. Luck has left us – the *p* value for the group comparison is now 0.065. The variability around the mean that we generated because we sampled too little obscures a difference that we know exists – the *CE*s are *not* "good enough".

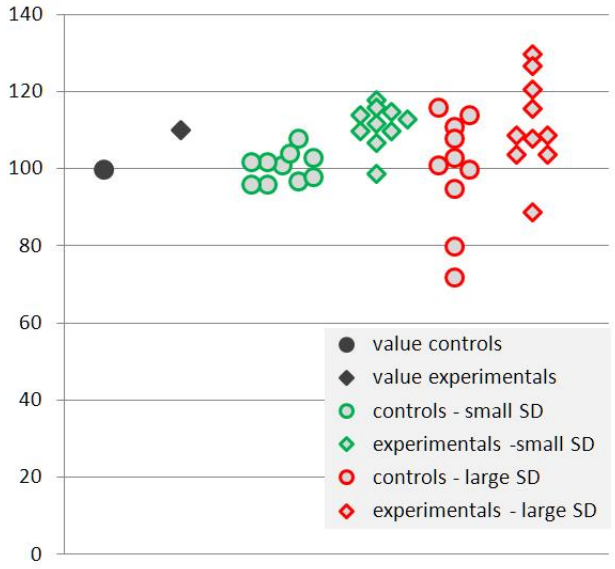

Figure 36 The impact of sampling generated variance on group variance. Sampling generates variability in a data set. Identical values (dark circle and diamond) may turn into a variable set of values with low (green) or high (red) variances depending on how much work we invest into sampling. If instead the green values reflect the true values of a parameter that has a natural variability indicated by the scatter of the point (circles and diamonds statistically different), adding variance by the sampling may turn them into the red data points (circles and diamonds statistically not different). Knowing the amount of variability that we add by sampling tells us if additional work could reduce the scatter in the red data points. It also tells us if we can expect a statistically significant outcome for the comparison between groups if we invest more work.

CEs are important because their size has direct influence on statistical outcomes. Drawing samples generates variability that adds to the natural variability present in the group. This relation is often expressed in the following simple formula.

$$Var_{group} = Var_{biology} + Var_{sampling}$$

Individuals in groups are, most likely, never absolutely identical. The true values may look like the green markers in Figure 36 reflecting biological variance ($Var_{biology}$). The variability added by the sampling ($Var_{sampling}$) may turn them into the red markers in Figure 36. Group variance (Var_{group}) is given to us by standard statistics. *A useful estimate of how much variability we added, i.e. the CE, would allow us to figure out if working harder on the sampling would allow us to obtain a statistically significant outcome.* If not, we need more individuals in our groups, i.e. a larger *n*. Not only that – if the statistical outcome of a group comparison has shown a significant effect, *a useful CE estimate would also allow us to figure out how much less work in terms of sampling or the number of individuals needed we may be able to afford with without jeopardizing our chances to see the effect*.

What *CE* value would be good enough for a descriptive study? In that the size of the CE has no immediate impact on further calculations the decision is somewhat arbitrary. One may aim for a *CE* that, because of its relatively small size, implies some quality. A value typically aimed for is a CE of ~0.1 or 10%. A more rational way is to adjust the CE that one aims for to the variability that is expected in the group (Section 9.8 and, e.g, West and Gundersen, 1990). One may also aim for a *CE* that in subsequent experimental studies involving statistical testing would allow the detection of group differences of a specified size for the least amount of work (Gundersen and Østerby, 1981). In case of the quantitatively nasty subventricular zone of mice (Azim et al., 2012), we found that decreasing the *CE* from 20% to 10% would require an increase in workload of 300% and allow the detection of a 20% effect and group sizes of five. Retaining a *CE* of 20% would require group sizes of eight, i.e. an increase in workload of "only" 60%, to detect a 20% effect. In this case, we decided that it would not be sensible to increase precision in a descriptive study because later experimental studies would more efficiently address their questions by a slight increase in *n* instead of aiming for a lower *CE*.

10.3 Calculating the CE based on a single sample

Several ways have been proposed to calculate the variance that originates from the sampling of sections. A prerequisite for their use is that the sections form a random systematic sample, i.e. that they form a series of sections of the type typically cut in a laboratory.

The key to understanding why the variability that originates from the sampling of sections can be calculated based on a single sample is to realize that counts obtained from one

section will always be able to forecast – more or less well – the counts that are likely to be obtained from nearby sections. The shape of brain structures was not generated by a random number table. If the cortex is large in section 10 and section 20 of a series, there is a good chance that it will be large in section 15 too (Gundersen and Jensen, 1987). Counts obtained from sections in a series co-vary with each other. The covariance between the successive values is captured in a covariogram, which can be approximated mathematically using the formula proposed by Gundersen and Jensen (1987) based on the work of Matheron (1965, 1971). For a volume estimate using counts of points in sections, P_i, the formula would be

$$CE = \frac{\sqrt{(3A + C - 4B)/12}}{\sum Pi}$$

in which

$\boldsymbol{A}$ is the sum, across all sections of the sample, of the counts in each individual section, i.e. P_i, squared

$\boldsymbol{B}$ is the sum, across all section of the sample, of P_i, multiplied by the counts in the following section of the sample, i.e. P_{i+1}, and

$\boldsymbol{C}$ is the sum, across all sections of the sample, of P_i multiplied by the counts obtained in the next to the following section, i.e. P_{i+2}

$$A = \sum_{i=0}^{n} {P_i}^2 \quad B = \sum_{i=0}^{n} P_i \times P_{i+1} \quad C = \sum_{i=0}^{n} P_i \times P_{i+2}$$

Note that the region, hopefully, was not present in the section that would be following the last section that was collected using the sampling scheme. The counts obtained from sections following the sample are zero, i.e. the last number to be summed up for *B* and the last two numbers to be summed up for *C* will be zero because P_i is multiplied by zero. Table 1 provides a brief example of the calculations using the small dataset generated in section *7.3 An example of a volume estimate*. The point count used in the example could be exchanged for any other count of probe-feature interactions.

Table 1 **An example of a CE calculation.** The example is based on the olfactory bulb point counts from Figure 13. Even though there are only seven sections in the sample in which only 70 counts were obtained, the CE generated by the sampling of the sections amounts to only 0.15 or 15% of the mean. The olfactory bulb was no longer present in sections 8 and 9. The counts are therefore 0, and the last entries in the two columns to the right will also be 0.

section i	P_i	$P_i \times P_i$	$P_i \times P_{i+1}$	$P_i \times P_{i+2}$
1	2	2 × 2 = 4	2 × 20 = 40	2 × 15 = 30
2	20	20 × 20 = 400	20 × 15 = 300	20 × 12 = 240
3	15	15 × 15 = 225	15 × 12 = 180	15 × 12 = 180
4	12	12 × 12 = 144	12 × 12 = 144	12 × 6 = 72
5	12	12 × 12 = 144	12 × 6 = 72	12 × 3 = 36
6	6	6 × 6 = 36	6 × 3 = 18	6 × 0 = 0
7	3	3 × 3 = 9	3 × 0 = 0	3 × 0 = 0
	ΣPi = 70	A= 962	B= 754	C= 558

$$CE = \frac{\sqrt{(3 \times 962 + 754 - 4 \times 558)/12}}{70} \cong 0.15$$

In this form, the Gundersen-Jensen *CE* estimator was used in, e.g., the paper that introduced the *Optical Fractionator* (West et al., 1991). Several improvements have been made since then (Gundersen et al., 1999) and will be introduced in the following sections.

Variability in the data obtained is generated each time a structure is sampled. Typically we sample a structure twice. The variance that originates from the first tier of sampling – when we decide in which sections of the structure the measurements are performed – is accounted for by the Gundersen-Jensen *CE* estimator in its original form. Variance originating from the second tier of sampling – when it is decided in which places within the section the probes area placed – can now also be accounted for. The way in which this component of the variance is estimated depends on the type of probe that is being applied. Fortunately, there are currently only two ways – one for the point probes used within sections to obtain a Cavalieri estimate of volume and a second one, for probes of area,

length and number that return a simple count of probe-feature interactions (e.g., isotropic lines, Spaceballs and Disectors).

A second area of improvement pertains to the *smoothness factor* of the structure that is being investigated. The ability of counts in one section to forecast counts in adjacent sections may differ between structures. Sections of a structure with a smooth surface (e.g., an almond) are more likely to resemble each other than sections of a structure with a crumpled surface (e.g., a walnut). How well they are forecast does, of course, also depend on the distance between the sections in the sample. The smoothness factor allows the adjustment of CE estimators to the quantitative morphological peculiarities of the region that we assess (see section 10.6 *The smoothness factor*)

10.4 Variance originating from volume estimator point counts in sections

This contribution to the variance of the estimate has unfortunately been referred to using a variety of name, e.g., local error, noise or nugget (referring to an irregular shape) variance. It is here termed S^2.

As already illustrated in Figure 11, the precision of an area estimate based on a point count in a section depends on density of the points used. Obviously, the point count will have to enter into the equation. Another factor that influences the precision of the estimate is the shape of the area. The counts that a very “thin” area will return when a point grid is applied to estimate area will be more variable than the counts that a very “thick” area may return. (Figure 37). Consequently, a second factor that needs to enter into the equation is the shape of the area to which we apply the points.

The relation between the variance that is generated by the estimate, point counts and shape was described by Matern (1985) and brought into an applicable form by Gundersen and Jensen (1987) by way of the following equation.

$$S^2 = 0.0724 \times \frac{b}{\sqrt{a}} \times \sqrt{n \times \Sigma P}$$

The point count obtained enters in the form of ΣP, i.e. the *total number of points* counted in *all sections*. Shape enters the formula in the form of the shape factor $\frac{b}{\sqrt{a}}$, i.e. the boundary length, b, of the structure divided by the square root of the area of the structure, a. This

factor grows when the region becomes thinner. *n* refers to the number of sections in which the counts were made.

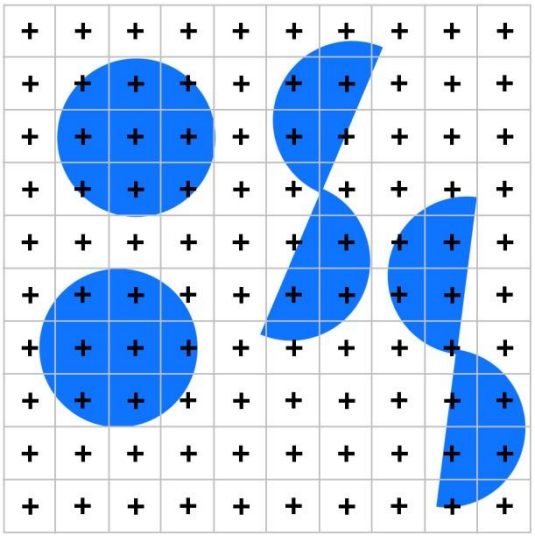

Figure 37 The impact of shape on the precision of point-count area estimates. Using this point grid, a circle with the smallest possible shape factor of 3.54 will, depending on how the grid was placed, return a count between 7 and 9. The second shape, composed of two hemicircles (shapefactor 5.8) with the same area of the circle, will return counts between 6 and 10. The top right corner of the crosshairs was used to count.

There is the problem of estimating the shape factor as both boundary length and area need to be known. While one obtains at least an estimate of the area from the point count itself, this is not so for the boundary length. Although it is fairly easy to estimate (Buffon, 1777; Cruz-Orive, 1997; Howard and Reed, 2010), one may decide not to bother. Even if the counts are done manually, points come very cheap in terms of the time needed to count them. If only 200 points are counted in 10 sections, the *CE* (now calculated only for the sampling within sections) for areas in the shape of a circle (the smallest possible shape factor of 3.54) would amount to only just about 1.7% of the area estimate.

$$\frac{\sqrt{S^2}}{\Sigma P} = \frac{\sqrt{0.0724 \times 3.54 \times \sqrt{10 \times 200}}}{200} = \frac{\sqrt{11.46}}{200} = 0.017$$

Very, very "thin" areas rarely exceed a shape factor of 30, and the variability would still amount to just less than 5%. Software packages allow thousands of points to be counted with ease, and 1000 points would limit the *CE* to 0.5% and 1.5% for shape factors of 3.54 and 30 respectively. Aside from this, software packages may even provide a rough approximation of the shape factor based on the boundary length of the area associated with the marked points. Point counting within sections will likely be a negligible source of variance if the number of points is as large as it easily can be. The section to section variability of the shape

of a region and, therefore, the sampling of sections will likely be the dominant source of variance.

For the olfactory bulb data in Table 1 and a shape factor of, e.g. 10 (close to that of hippocampal cell layers and certainly larger than the actual one),

$$S^2 = 0.0724 \times 10 \times \sqrt{7 \times 70} \cong 16$$

or a contribution to the *CE* of $\frac{\sqrt{S^2}}{\sum P} = \frac{\sqrt{16}}{70} \cong 0.06$ or 6% - not bad for counting a total of only 70 points in seven sections.

10.5 Variance originating from counts using number, length or surface estimators in sections

While an understanding of the variance generated by point counts is rather intuitive – many points ≈ lots of precision, crumpled shape ≈ less precision – this is not so for the variance originating from other probe-feature interactions in sections. Currently, there is no statistical well-founded approach to estimate it. Instead one has to resort to a non-intuitive approximation that, at least, has the advantage of extremely easy calculation.

$$S^2 = \Sigma\, interactions$$

If 100 cells are being counted in Disector probes, S^2 for the number estimate would be 100. If we observe 214 interactions of capillaries with the surface of spaceballs, S^2 of the length estimate would be 214.

The assumption that allows this easy calculation is that the number of interactions that we count each time we place a probe originates from a Poisson process (no panic! Cruz-Orive and Geiser, 2004). The usefulness of the assumption has been verified in models (Schmitz, 1998) and real object populations, in which the Poisson approach provided a conservative estimate (Cruz-Orive and Geiser, 2004). A Poisson distribution has some convenient mathematical properties. The variance of a Poisson distribution is equal to its mean, and the sum of two Poisson distributions is a Poisson distribution. A Poisson distribution has been constructed to behave like that mathematically. For us, there is no real reason to understand why it behaves this way, but we need to know, for just a few seconds longer, *that* it behaves this way.

If the count coming from one probe originates from a Poisson process and the count coming from another probe does so to, then the sum of the two counts from the two probes comes also from a Poisson process. Continuing like this for all the probes applied across all sections in an individual, the sum of the counts from all probes also comes from a Poisson process. How many times did we apply *all* probes? Once. The mean of our sum is therefore sum/1. What we want to know is not the mean but the variance, which, if we are looking at a Poisson distribution, has the same value as the mean. Voilà – $S^2 = \Sigma\ interactions$.

Probe-feature interactions for number, length or area are harder to come by than point counts for volume. Routinely, one is first roughly outlining the structure of interest, which takes a non-trivial amount of time even if done coarsely. Subsequent work with high magnification oil lenses requires time to change lens, oil or clean section and step across the tissue, and we of course need to look for the probe-feature interactions. Counting in excess of 200 interactions quickly becomes tedious work. And 200 interactions could still generate a 7% error.

$$\frac{\sqrt{200}}{200} = 0.07$$

In contrast to point counts, variance originating from sampling within sections with probes for number, length or area may contribute a large part to the total variance that we introduce with the sampling.

10.6 The smoothness factor

As mentioned above, the ability to forecast counts in sections based on the counts in close-by section depends on both the shape of the structure that is being assessed and the distance between the sections. If the distribution of the feature of interest changes very gradually from section to section, even sections that are spaced far apart may provide a good forecast. If the structure is shaped very irregularly, even closely spaced section may have a poor ability to provide a good forecast of what is happening in adjacent sections. It is *not* the shape of the structure itself that needs to be known to make the forecast, but just the counts obtained from the sections that were sampled.

The quality of the forecast that can be expected from a sample of sections enters the calculations of the *CE* by way of the *smoothness*. Smoothness refers here to the appearance

of the curve that is formed when the data points obtained from the sample are plotted in order. We need, of course, some number that can enter into the calculation of estimate precision. The smoothness of the curve can itself be estimated (Cruz-Orive, 1999; Cruz-Orive, 2006; Gundersen et al., 1999; Kiêu, 1997), and stereology software packages may include an estimate of smoothness in their output. Estimators of smoothness are mathematically unbiased, but they are not robust when based on the datasets that are available from typical quantitative studies (Cruz-Orive, 1999; García-Fiñana and Cruz-Orive, 2004; Gundersen et al., 1999), i.e. individual estimates may be far off the true value. Also, estimates of smoothness must be mathematically converted into the smoothness factor, $\alpha(m)$ or just α, before it can enter into the equation that will provide the estimate of the CE. This conversion does still require the use of interpolation tables (Cruz-Orive, 2006). The bottom line is that smoothness is currently judged rather than calculated – perhaps also because estimating something to enter into an estimator of something related the variability of an estimate is exhausting the trust in the number that is finally generated.

The smoothness, *m*, can take any value equal to or larger than zero. The higher the value of *m*, the smoother is the appearance of the plotted distribution. Typically only two values of *m* are considered – zero or one. *m* can be set to one if there are all jumps (changes significantly exceeding the point-to-point jitter) can be "predicted" by the preceding or succeeding points (Figure 38A). *m* can be set to zero if the distribution contains sudden jumps that are not "predicted" by the preceding or succeeding points (Figure 38B). If one is conservative and/or uncomfortable with making this judgment, one may decide on *m* being zero independent of the appearance of the distribution. Estimates of the precision are unlikely to be worse than the ones obtained for an *m* of zero (Cruz-Orive, 1999; Cruz-Orive, 2004).

Note that while selecting the correct *m* does impact on the quality of the *CE* estimate, an *m* of zero does not generate *CE*s that necessarily must be much larger than those predicted by an *m* of one. Samples comprised of twelve sections will generate CEs between only two and three percent for both structures illustrated in Figure 38 (Basler, 2011; Slomianka and West, 2005), even though they are best predicted by a different *m*. That one may expect small CEs from both samples is also suggested by the good fit between the lines connecting the data points of the sections belonging to the samples and the underlying volume distribution.

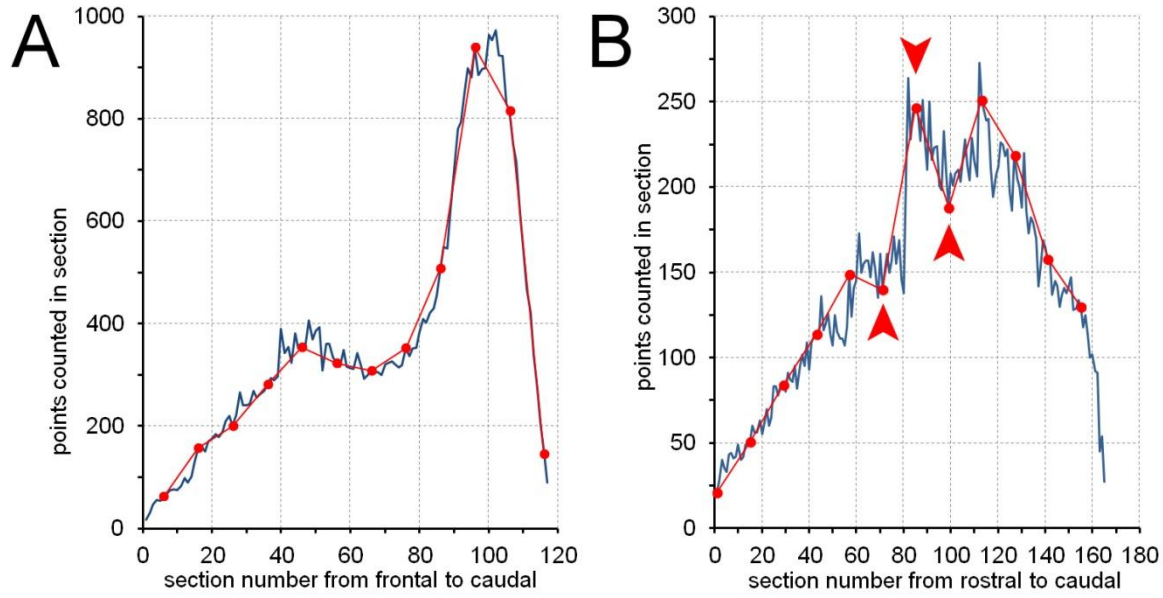

Figure 38 The smoothness of volume distributions. The graphs show samples of 12 sections (red circles) of (**A**) the entire mouse dentate gyrus and (**B**) the rat hippocampal CA1 pyramidal cell layer in coronal series of sections. Dark lines show the volume distribution based on all sections. Sample values in (**A**) appear reasonably predictable by the surrounding values and an *m* of one seems suitable for this sample, which also has been verified empirically (Basler, 2011). The distribution in (**B**) does show sudden jumps (arrows) that do not appear to be predicted by the preceding or following data points. The appearance of the data plot and empirical results (Slomianka and West, 2005) suggest that an *m* of zero would be the most suitable choice.

For the two cases of *m*, we can finally obtain smoothness factors to enter into the CE calculations.

*For **m = 0**, the smoothness factor, α, is **1/12**; for **m = 1**, α is **1/240**.*

The value to be used for an *m* of zero did already appear in the equation introduced in section 10.3 *Calculating the CE based on a single sample*. We now have the opportunity to change it according to our perception of the smoothness of the count distribution that we obtained from a sample.

$$for\ m = 0 \quad CE = \frac{\sqrt{(3A + C - 4B)/\mathbf{12}}}{\sum Pi} \qquad for\ m = 1 \quad CE = \frac{\sqrt{(3A + C - 4B)/\mathbf{240}}}{\sum Pi}$$

With the improvement of the *CE* estimator, it was suggested that an *m* of 1 would be a more satisfying approximation than an *m* of 0 (Gundersen et al., 1999). Software packages may provide *CE* estimates for *m* of both zero and one. These values may be taken as the upper

and lower limits (e.g., Filice et al., 2016), which, independent of the spacing of the sections, are rarely exceeded (Basler, 2011; Cruz-Orive, 2004; Slomianka and West, 2005).

10.7 The current Gundersen-Jensen *CE* estimator

In the original form, the Gundersen-Jensen estimator did not account for the fact that the counts obtained from the sections are themselves only estimates. The count that is used for a section may not be small because the section is small but because, by chance, there were few probe-feature interactions despite the fact that there were ample of probes. Because we only have estimates for the true values of the sections, we underestimate the ability of the sections to forecast counts in nearby sections and we therefore overestimate the variance that is generated. In sections 10.4 and 10.5, estimates of the variance, S^2, originating from the probe-feature interactions in sections were calculated. It is subtracted in the term that defines the variance originating from the sampling of sections, because it led us to overestimate the error produced during the sampling of sections. On the other hand, it is added to the total variance as an independent component, because we actually generate this variance during the sampling within sections.

Together with the possibility to adjust for the smoothness that we perceive, we now have the G-J *CE* estimator in a generalized form that is in use today.

$$CE = \frac{\sqrt{(3(A - S^2) + C - 4B) \times \alpha + S^2}}{\sum f}$$

in which

$\boldsymbol{S^2}$ is the variance originating from the sampling within sections – calculated either according to the formula given for point counts in the Cavalieri estimator in section 10.4 or according to the formula for probe-feature interactions using number, length or surface estimators in section 10.5.

$\boldsymbol{f}$ is the value of our measurement function – a term that covers any type of counts obtained in a section using the estimators presented in this thesis. Σf is the sum of the counts of whatever across all sections.

The *CE* of the olfactory bulb data set (Table 1) can now be recalculated, for the last time, using the current Gundersen-Jensen estimator. A *CE* estimate of ~0.16 is obtained for a

smoothness factor (*m*) of 0 and an estimate of ~0.07 for an *m* of 1. Had we counted 70 cells instead of 70 points, the *CE* estimates would be ~0.19 for an *m* of 0 and ~0.12 for an *m* of 1.

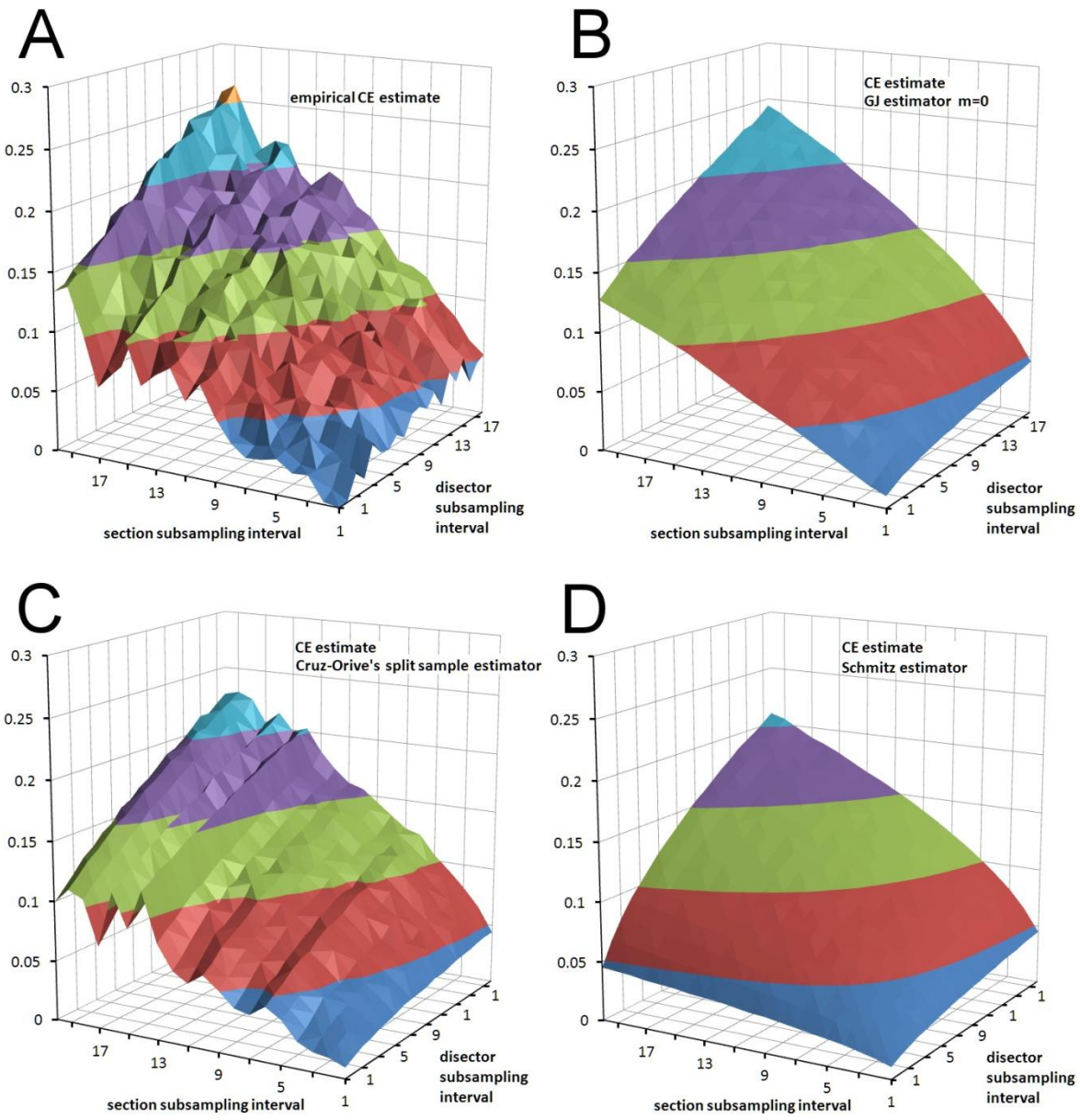

Figure 39 **Empirical CE estimates and CE estimators**. In (**A**), CEs of a number estimate were estimated empirically by collecting a very large data set () from hippocampal CA1 pyramidal cells. From this data set subsamples were drawn and the CEs were estimated (CV of the mean of all subsamples for each combination of subsampling intervals). The Gundersen-Jensen (GJ) estimator (**B**) and Curz-Orive's split sample estimator (**C**) provide useful approximations of the variance generated by the sampling. The estimator of Schmitz (1998) (**D**) considers only the variance component generate by sampling within sections and provides a simple to calculate, rough-and-ready estimate *if* the sampling of sections contributes very little variance to estimates.

Although trying to convey an intuitive understanding of how this estimator works fills some space, and although the formula looks fairly impressive to non-mathematicians, the most complicated mathematical operation that has to be performed is drawing a square root. A *CE* estimator is easily implemented in a spreadsheet, in which one only would have to enter the counts obtained from the sample of sections in their correct anatomical order to obtain a *CE* estimate. I have even seen *CE*s being calculated using mobile phones.

What we gain from either the small effort that is required to generate a *CE* estimate or from "just" understanding the estimates provided by software packages is a way to judge if the estimates that were performed are not only statistically valid – they always will be provided that the sampling was representative – but, finally, also if they are "good enough" in the context of the study in which the estimates are used.

10.8 Finally: Good enough or not?

We have now two of the variables that formed the equation presented in section *10.2 Why is a CE important?*.

$$CV^2_{group} = CV^2_{biology} + mean\ CE^2_{sampling}$$

The relative group variance, i.e. the standard deviation squared divided by the mean, can be calculated from the mean and standard deviation provided by routine statistics – whether they are part of a pocket or desktop calculator, a spreadsheet or more or less advanced statistics packages. It would be difficult to avoid seeing them somewhere if one tried. We do now have a means to calculate the *CE*s. Note that the mean CE^2 is calculated as

$$(CE^2_{animal\ 1} + CE^2_{animal\ 2} + \cdots + CE^2_{animal\ n})/n$$

and not by first calculating a mean *CE* and subsequently squaring it.

Well, when is a *CE* good enough? It is good enough if it is not the weakest link of our quantitative procedures. If something needs to be improved, the proper place to invest work would be the weakest link. The ratio between the mean of the CE^2s and the relative group variance will provide a guess at where the weakest link is located (Gundersen and Østerby, 1981).

$$\frac{mean\ CE^2_{sampling}}{CV^2_{group}}$$

If this ratio is smaller than 0.5, the sampling introduced variability contributes less than half of the total variability that is seen between animals. Natural differences between animals contribute more to the variability seen in the group than the sampling that we used. If we wanted to improve the reliability of the means presented in a descriptive quantitative study or if we wanted to improve the chance of statistically detecting a group difference, the proper place to invest work would be the source of most of the variance. If the main source is natural differences between animals, increasing the number of subjects would be more efficient than increasing the amount of work that we invest into each subject. The *CE*s are good enough; if necessary, we ought to increase *n*. If the ratio is larger than 0.5 the reverse applies. Sampling is a larger source of variance than natural differences between animals. Improving the data by investing more work into the subjects at hand would be more efficient than increasing the number of subjects. Strictly speaking, the *CE*s are not good enough.

How does one handle *CE*s that are not good enough? It depends on the outcomes of statistical testing, the workload associated with increasing *n* or improving *CE*s, but, first, a closer look at the available data. In, e.g., Amrein et al (2004), we reported a ratio of 1.44 for one of the species in which hippocampal granule cells were counted. The estimate of the sampling-induced variability by far exceeded the variability that was observed in the group. This ought to be impossible – if we introduce variability by estimating, we ought to see it in the group. The ratio between what we introduce and what we see should never be larger than 1. Yet it is possible because we talk statistics. It is possible because both the estimates of the *CE* and the observed group variance are just that – statistical *estimates* based on samples. We may, by chance, have drawn a sample that generates a *CE* estimates that is larger than the real *CE*. We also may, by chance, have drawn a sample of animals that are very similar to each other. In this particular case, the *CE*s for the granule cell counts were less than 0.1 – the value that we also aimed for. The group *CV* for the granule cell counts was however as low as ~2%. It is not clear if this was chance or if granule cell number is very tightly regulated in the species. It is clear however that we would still be counting today if we had attempted to decrease *CE*s to a value that would return a ratio below 0.5. Despite the large ratio, we considered the *CE*s to be good enough.

As also the case in our study, statistical outcomes by themselves may, for the time being, justify a poor CE^2/CV^2 ratio. Sampling may be the weakest link, but if one actually has detected a significant group difference, sampling was apparently not sloppy enough to hide a significant effect. Note that it also means that a smaller *n*, which is more responsible with regard to animal experimentation ethics, would be sufficient to detect the difference if the sampling is improved.

If assessments of the data and/or statistical outcomes do not provide a loophole for poor ratios, it is time to think. A ratio below 0.5 is advice towards more subjects. Above 0.5, the advice is more sampling. Whether one follows the advice depends on the workload and feasibility associated with increasing *n* or decreasing the *CE*. Variance may need to be just a little better to generate an outcome; additional subjects may not be available but additional sections may still be available from the subjects that we have. In this case it may be worthwhile trying to improve estimates even though it would not be the most efficient way to obtain an outcome. A fully calculated example that can provide a clue if it is even worth trying can be found in West (2012c).

10.9 Other CE estimators

Some emphasis has been put on the Gundersen-Jensen estimator because it, for the structures that we have been working with, delivers estimates that are close to the *CV*s that can be observed empirically (Basler, 2011; Slomianka and West, 2005) and because it is the estimator most commonly used. Several other CE estimators are available. A particularly attractive one is the Split-Sample estimator introduced by Cruz-Orive (Cruz-Orive, 1990; Cruz-Orive and Geiser, 2004). It is based on the mathematical ideas also at the root of the G-J estimator. However, the formula to calculate it looks quite different. The Split-Sample estimator is intuitively easier to understand than the Gundersen-Jensen estimator. We can split our data in half. If the estimates based on only half the data are very similar to each other and therefore also to the estimate based on all data, than the estimate based on all data must be pretty robust. If doing less would not have made the estimate much less reliable, doing more probably would not make it much more reliable. Of course, a *CE* can be calculated.

$$CE = \sqrt[2]{\frac{(1-\tau)^2}{(3-2\tau)} \times \left[\left(\frac{Q_0^- - Q_e^-}{Q_0^- + Q_e^-}\right)^2 - \frac{1}{Q_0^- + Q_e^-}\right] + \frac{1}{Q_0^- + Q_e^-}}$$

in which τ is the sampling fraction and Q_o^- and Q_e^- are the counts obtained from the odd and even sections of the sample.

Note that the formula also will return a *CE* estimate of zero if we count everything that there is, i.e., if the sampling fraction, τ, is 1. Similar to the Gundersen-Jensen estimator, the Split-Sample estimator return *CE* values that correspond well to values observed empirically (Slomianka and West, 2005) but without a need to find a smoothness factor or to calculate S^2. Lastly, *CE* estimates returned by this estimator do not increase smoothly but show regional troughs and peaks that are very similar to the *CE* landscape observed in the empirical data (Figure 39C).

Using computer simulations, both Glaser and Wilson (1998) and Schmitz (1998) found that $1/\sqrt{\sum Q^-}$ closely approximated the empirically observed *CE* in their models. Note that this approximation is identical to a GJ *CE* estimate based on S^2 alone. Notably, section to section variability was non-existent or minimal in these models. They therefore confirm the validity of S^2 and that the *CE* will be dominated by S^2 if regions are very smooth. *If* a region is very smooth, one may use a CE based on S^2 alone as a rough-and-ready GJ *CE* preview.

10.10 Estimate precision and the orientation of the sections

Estimate precision depends on how well a sample captures changes in the distribution of the structures of interest along the axis of sectioning and in their distribution within sections. Changing the orientation in which the sections are prepared often changes the appearance of the region that contains the structures of interest. If it does, it may also change the distribution along the axis of sectioning. If the distributions differ in their smoothness there will also be differences in the efficiency of the sampling. One of the three directions in which the central nervous system is usually cut may return a much higher precision for a given amount of sampling than the other two directions. If workload is a critical factor, cutting directions may be evaluated for their suitability to sampling before a final decision on how to cut the tissue is made.

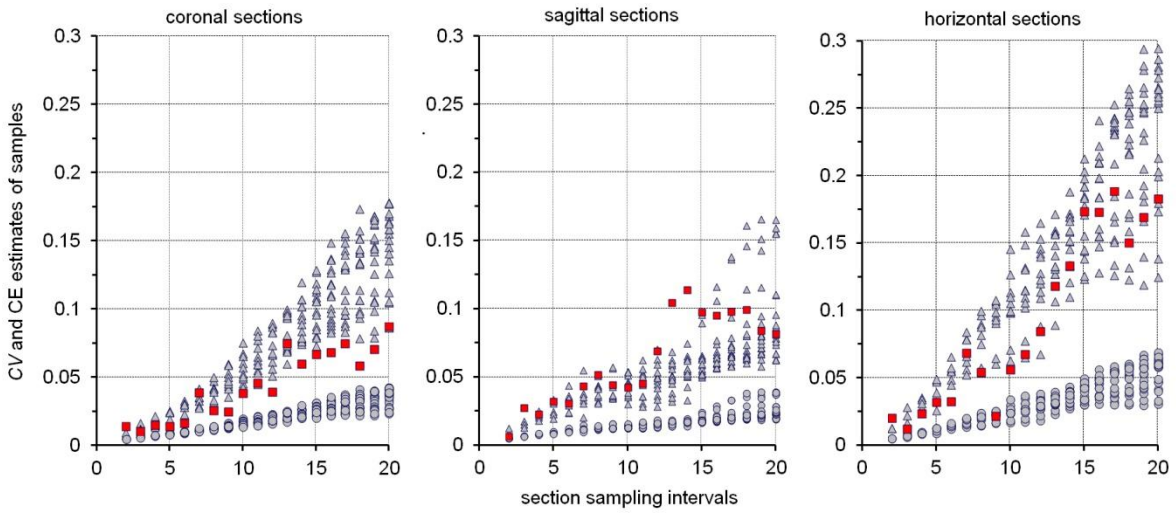

Figure 40 **Sectioning and estimate precision.** Graphs show the empirically derived CVs (red) and GJ CE estimates (triangles for m=0 and circles for m=1) for the mouse dentate gyrus granule cell layer (data from Basler, 2011). CV estimates are based on exhaustive series of 20 µm thick methacrylate sections. The graphs represent the CVs and *CE*s for a coronal, a sagittal and a horizontal series of sections. With larger intervals between sampled section, *CE*s increase slower when the coronal and sagittal series are sampled than when the horizontal series is sampled. With the exception of sampling intervals between 12 and 15 (see section *10.11 CE estimators and systematic variations in morphology*), empirical *CV*s are estimated reasonably well by *CE* estimates using an *m* of zero.

Figure 40 illustrates the effect of sectioning direction on the efficiency of sampling for the dentate gyrus granule cell layer in the mouse hippocampus. First, the cell layer requires about 110 sections (20 µm thick) to be cut when the brain is sectioned coronally (Figure 35) or sagittally but close to 140 horizontal sections. If the sectioning is restricted to this region of interest, workload for the collection and processing of section can be cut by ~20% by deciding in favor of coronal or sagittal sections. Coronal sections return a *CE* consistently below 5% for the sampling intervals up to 12 and well below 10% for all tested sampling intervals up to 20. Sagittal sections are only slightly less efficient, but precision is less reliably predicted (see section *10.11 CE estimators and systematic variations in morphology*). Horizontal sections do not only require more sections to cut, but precision decreases more rapidly with increasing sampling intervals. The *CE* for larger sampling intervals exceeds that of coronal or sagittal section by almost a factor 2. The largest tested interval, i.e. every 20th section or 5 to 6 section per granule cell layer (110/20), returns a CE of 8% from coronal sections. To obtain the same precision in horizontal sections, every 12th section or ~12 sections (140/12) have to be examined. In terms of the sections that need to be prepared

and analyzed, coronal sections actually require half the work to obtain a precision of 8%. Savings of 20% while cutting and 50% while preparing and analyzing sections are well worth considering in large scale projects. Similar differences in sampling efficiency between sections of different orientation were also observed for the hippocampal pyramidal cell layer (Slomianka and West, 2005).

Alas, our laboratory often uses horizontal section for this region of the brain because the surrounding cortical areas are much easier to assess using this orientation. Efficiency is a factor, but may only be one of many others that enter into the decision making process.

10.11 *CE* estimators and systematic variations in morphology

What happens if there are systematic changes in the anatomy of the region of interest? A slightly malicious glee that sometimes accompanies this question hints at the discovery of a major flaw in the procedures suggested for sampling and the assessment of estimate precision. For many regions of the brain we know that they are composed of repetitive units with distinct anatomical appearances that represent functional entities – cortical columns or barrels, thalamic barreloids or cerebellar aldolase stripes to mention but a few. If the distance between the samples matches the distance between the anatomical units, not every part of the unit will have a chance to be included in the sample. Consequently, sampling as outlined in section *5.4 Fractionator sampling* would not be representative and, consequently, the estimate generated and the assessment of sampling precision would be fatally flawed. This supposition is as easily dismissed as it is to step into the actual traps that anatomy has lain.

The supposition *may* turn out to be true if (1) there is a near perfect match of sampling intervals and the size of the anatomical unit across the entire section *and* (2) if very few sections are used *and* (3) if there are very few animals per group. As soon as the intervals do not match perfectly, the units will, sooner or later, again be hit in different places. E.g., cortical curvature alone will change the spatial relation between cortical columns and a rectangular sampling grid. The section or the sampling grid may also be slightly rotated to minimize the chances that intervals stay synchronized – an option that is part of stereology software packages. If multiple sections are used and a random starting point is used in each section, the units will be hit in different places in different sections even in the unlikely event that this does not occur in each section. In the very unlikely event that units are hit in the

same place in each section and across multiple sections, there is still the chance that they are hit in different places in different animals. If we consistently, against all statistical odds, hit the same spots, the estimate obtained would still be statistically valid, because it is the outcome, however statistically unlikely, of representative sampling and probing. The estimate would, of course, be practically useless. Notably, the chance of this happening is only smaller but still present if random independent samples are used. Suffice is to say that sampling intervals should avoid known anatomical regularities to keep the good odds as good as possible. Problems of periodicity when sampling in sections could be avoided by the selection of unaligned sampling (e.g., Figure 6 in Cruz-Orive and Weibel, 1981), in which the sampling locations still sample a systematic set of areas defined by the x- and y-steps, but they are placed at a random location in these areas. Unfortunately, unaligned sampling has not been implemented in the major stereology software packages.

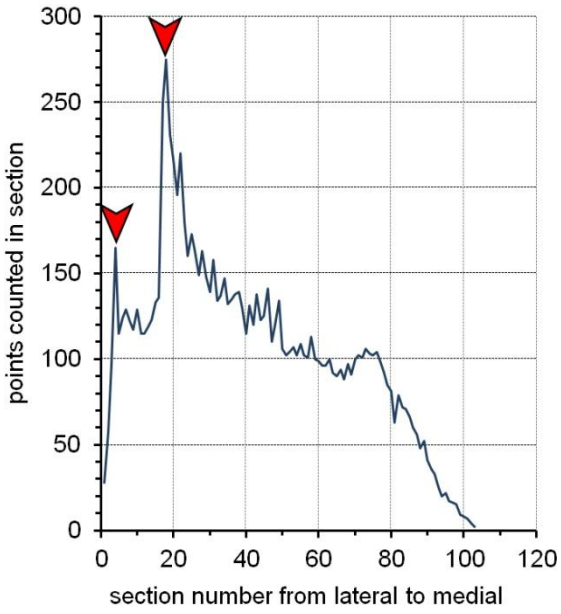

Figure 41 **Periodicity in a volume distribution.** Sagittal sections of the mouse dentate granule cell layer will twice pass almost horizontally through the layer. The large areas of the layer in these sections will result in large point counts (red arrows). If a sample of sections include or miss both peaks (section sampling intervals of 13 or 14) the resulting volume estimates will very large orvery small generating the large *CV*s for these intervals in sagittal sections shown in Figure 40 (data from Basler, 2011).

Rather than the functional units of the brain, it is the unexpected quirks of anatomy that may play tricks with our ability to obtain precise estimates and to formally assess precision. In Figure 40, the empirical estimate of the *CE*s obtained from sagittal sections do, for a short

distance from sampling intervals 13 and 14 exceed any of the estimates that are provided by the Gundersen-Jensen *CE* estimator and an *m* of zero. Looking at the volume distribution of the sagittally sectioned series in Figure 41, we see two sharp peeks that are separated by a distance of 13 to 14 sections. With these sampling intervals, either *both* peaks are contained within the samples of sections or *both* peaks are missing from the samples, which generate either very large or very small estimates. The variance generated by the sampling is not detected by the Gundersen-Jensen *CE* estimators, which may underestimate actual sampling related variance for these section sampling intervals by a factor 2. It is the folding of the cell layer in relation to the plane of the section that plays a trick on the *CE* estimates. Sagittal sections will twice pass parallel trough a large part of the cell layer, and two peaks are sufficient to generate a period. That does not make *CE* estimators useless. It does however mean that having a good feeling for how anatomy may impact on the precision of estimators is helpful when evaluating estimates of precision. Once again, sampling intervals close to such periods should be avoided.

10.12 Designing a useful sampling scheme

For the sampling scheme, we need to select parameters for the selection of the sections, for the selection of sampling locations within each section and for the size of the probes that will be applied at each sampling location.

Selection of the sections – Selecting sections means deciding on the spacing of the sections, i.e. the section sampling fraction, in the series that will be used for an estimate. The number of the sections that are needed will depend on the shape of the region that is to be investigated and the distribution of the structures of interest within the region. Shape and distribution define how much signal there is in each section. If a region of interest appears large in a section, we can expect to see many of the structures of interest if they are more or less evenly distributed in the region. If the structures are distributed unevenly, the amount of signal depends on both the size of the region in the section and the local density of the structures. Key to the selection of a useful section sampling fraction is not to miss the parts of the region which contain most of the signal that we are interested in. Looking at the volume distributions of hippocampal divisions and cell layers in Figure 35 and Figure 38, I would not want to miss the large peaks. This means looking at the width of the peaks, and deciding on section sampling fractions that place two or three sections within the peaks. If

the signal is distributed rather evenly, the use of about ten or even fewer sections has been suggested (Gundersen et al., 1999). A rather even distribution of the signal would mean a smoothness factor close to 1 and, using this number of sections, only a small contribution of section-to-section variability to the total variability of the estimate.

The final decision on the section sampling fraction must rest with the person most experienced with the distribution of the signal in a region – more likely the principal investigator rather than a stereologist knowing not much more about a region than its name.

Selection of sampling locations – In section 5.3 *Statistically representative sites within sections*, it was shown that any unevenness in the distribution of the structures of interest will be captured better and better the more sampling sites are placed in a region. A typical recommendation would be about 100 to 200 sampling sites (Gundersen et al., 1999). Next, the distances between sampling sites along the x- and y-axes of the sections need to be estimated. This requires an estimate of the area of the region of interest that is available in the sections selected for analysis

In section 7.3 *An example of a volume estimate*, point counts were used to estimate the area of combined granule cell, internal plexiform and mitral cell layers in the hamster olfactory bulb. The total number of points counted was 70 and the area associated with each point was 125,000 μm^2. The total area in the sections is therefore 70 × 125,000 μm^2 = 8,750,000 μm^2. If the region is to be hit by 100 probes, it needs to be hit once for each $1/100^{th}$ of its area, i.e. once for each 8,750,000 $\mu m^2/100$ = 87,500 μm^2. The step size along the x- and y-axes would therefore be the square root of 87,500 μm^2 or ~296 µm. If the estimate was made in a series and the decision is made to rather pool two series for actual quantification, about twice the area will be available. Step sizes of ~417 µm would return ~100 sampling sites.

As already pointed out in section 7.3 *An example of a volume estimate*, the area sampled for each step along the x- and y-axis does not need to be a square. Depending on the looks of the region in a typical section a rectangle may be more suitable. E.g., if a region forms a more or less vertical, long and narrow band, it may be a good idea to select smaller x-steps and larger y-steps (West and Gundersen, 1990). This increases the chances to occasionally hit the region in each section instead of hitting it very often in some sections and very rarely

in others. The variability of the estimates from section to section and, thereby, the variability of the final estimate, will decrease.

Also, 100 sampling sites should be amended with "of which a satisfying fraction returns some signal". Structures of interest may be heavily clustered within the sections, e.g., all the structures of interest may be focused within *one* spot about the size of $1/100^{th}$ of the total area. Of the 100 samples applied, there would be 99 empty samples and only one which contains signal. It also means that the cluster may be completely missed in some animals while being hit twice in other animals. This would massively increase estimate variability. If there are many clusters within each animal the danger of hitting very few or very many decreases. Depending on the perceived heterogeneity of the signal distribution step sizes may be adjusted to increase the number of samples that return a signal.

Selection of the probe size – If a site of the section is hit that contains structures of interest, the probe should be sized to also return a signal from this sampling site. We should, on average, observe one or more probe-feature interactions. A Disector should, on average, return a count of one or more cells, or a spaceball should, on average, intersect one or more axons. There will always be some variability that depends on the distribution of the structures of interest within the region, and some probes may well return a count of zero. Without before-hand knowledge, selecting the probe size is a trial-and-error process. One or more sections are selected and a probe is tried out in a few sampling locations chosen more or less at random. If there are much more than two counts with each probe and no or very few probes that do not return a count, the size could be decreased. If there are many probes that do not return a count and few that return a count of one or two, probe size should be increased. Instead of using probes that consistently return high counts, it would make more sense to use a smaller probe and more sampling sites. First, more sampling sites would allow a better coverage of any unevenness in the distribution of the structures of interest and, second, it may be visually more comfortable to count few interactions with each probe. While illustrative of many cases related to counting rules, most of the figures of probes in this thesis show more probe-feature interactions than one would hope to see in an average probe used to actually count.

With about 100 sampling sites and probes that return between one to two counts on average, the final count will be in the range of 100 to 200. Note that because of S^2, the CE obtained with the Gundersen-Jensen estimator cannot be smaller than the square root of

the count divided by the count. For a count of 100 that would be 10/100 or 0.1 or 10%. Improving this number would demand significant amounts of additional work. To decrease the *CE* to 5% we would have to obtain a count of 400. An overall *CE* of 10% is often aimed for (e.g., Mills Schumann and Amaral, 2005) because it can be achieved with a reasonable amount of work and, maybe more importantly, because in many cases a *CE* of this size accounts for little of the group variance (Gundersen and Østerby, 1981). Group variance is therefore more efficiently improved by adding animals instead of counting more within each animal.

Better be safe than sorry – When a sampling scheme is designed, one may add a little to everything – maybe a few more sections than one expects to need, some extra sampling sites and a slightly larger probe. There are three reasons to do so.

First, it is not known if the animal that was used to design the sampling scheme is representative of the group, i.e. close to the group average. If the animal would return large counts compared to others in the group, a sampling scheme that keeps to minimum requirements will return few and maybe too few counts in the other animals.

Second, even if the animal would be close to the average of the group, the sample that we drew from this animal may not be close to the average of this animal. The series of sections may have contained one more section or a larger section than other series. The area estimate may have been on the large side of the average, and the probe may have accidentally returned a few more cells from the sites at which it was tried than it would do on average. Again, cutting it close may result in subsequent estimates to return less signal than it was hoped for. At this point, e.g., the step size that has been calculated may be rounded down to a number that is both easy to remember and to report. Instead of the 296 µm that were calculated above, one may use 280 µm, which should increase the number of sampling sites by about 10% ($296^2/280^2$).

Lastly, nothing is more frustrating than having to do the same animal twice. Being generous with parameters at the first trial does not only mean that there are solid data to adjust the sampling scheme. It also means that one is definitely done with the animal.

Assessing the sampling scheme – Once the first complete estimate has been generated the sampling scheme should be revisited. Was the desired number of sampling sites obtained using the selected step sizes? Did the probes generate the number of counts that was aimed

for? Which *CE* was obtained using the sampling scheme? According to the answers to these questions the step or probe sizes or the number of sections used can be increased or decreased. After the next estimate has been generated with the revised sampling scheme it can be revisited again.

11 Quantitative Morphology at the Bench

The beauty of design-based stereological method lies in their very strong mathematical foundations. If these methods are applied correctly and within the limits of what is technically possible, we will obtain a *correct number* associated with *the structures that we have probed in our samples of the tissue*. This is a fact that is about as futile to discuss as the fact that one plus one equals two. Another fact that hardly needs discussion is that whenever a beautiful theory is faced with everyday life there will be problems, and when design-based stereological methods are faced with the lab bench there will be problems too. Guard zones have already been introduced in the context of the optical Disector probe. Mathematical theory does not require the use of guard zones. However, artifacts that could limit our ability to recognize objects may be present at the surfaces of sections. Such artifacts may make guard zones a necessary evil. Guard zones represent purely practical adjustments to the reality of the sectioned material and the abilities of the observer. While guard zones have become built-in parts of methods, there are other sources of possible error that do not have built-in safeguards. In fact, "*the structures that we have probed in our samples of the tissue"* is a qualification to "*correct number*" that leaves plenty of room for questioning the numbers. Are all structures that we should be interested visible in the sections? How does tissue processing – before we ever have a chance to look at a sample – influence the parameter that we are interested in? And are the representative samples that we draw as representative as they ought to be? Some of the problems that need to be solved to answer these questions positively are addressed in the following sections.

Note that these are purely practical problems that have no bearing on the theoretical validity of stereological methods. More importantly, note that any quantitative morphological approach is likely to be faced with one or more or all of the very same problems. However miserable we will fail in addressing these problems, we will fail least miserable if the methods that we use are not fraught with intrinsic problems and, at least in theory, must provide the right result.

11.1 Measuring section thickness

Optical Fractionator estimates typically require measurements of section thickness. Also, a volume estimate may require a measurement of section thickness if the estimate is used as a reference volume for estimators that themselves only return density estimates. Currently there are no methods to generate an unbiased estimate of section thickness. In contrast to the Disector probes, in which errors at the top and bottom faces of the probe cancel out, errors in measuring section thickness are likely to be additive (Figure 42; West et al., 1991). A measurement of section thickness requires the definition of the top and bottom of a section to determine the distance between them. This in turn requires decisions about when features at the top and bottom appear in the focal plane of the objective lens. The focal plane actually has a thickness – the "depth of field". Features at the very top and bottom of the sections will be in focus at all possible locations within the depth of field (Figure 42), and an observer may choose any of these positions depending on the perception of focus.

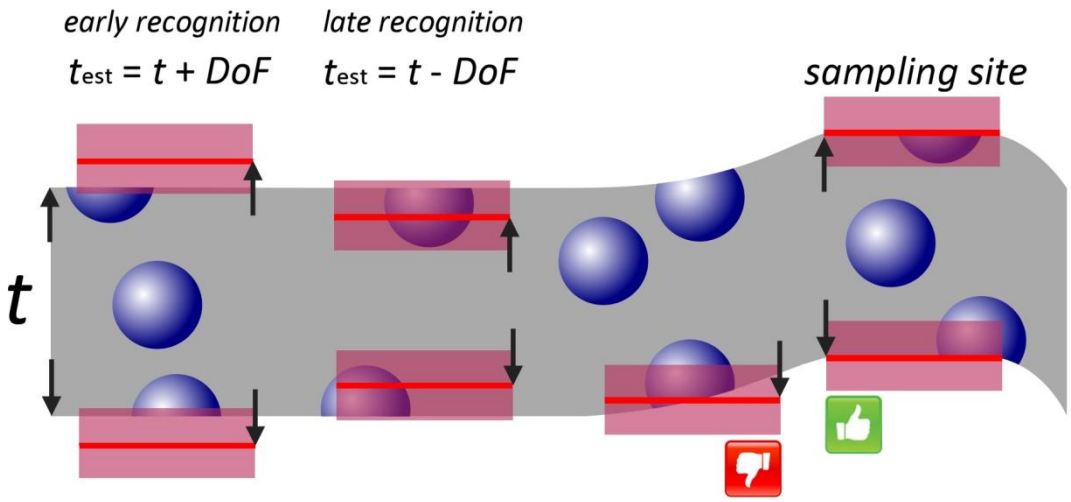

Figure 42 **Measuring section thickness.** Section thickness, *t*, can only be measured with a precision to within twice the depth of field (*DoF*, red boxes) of the objective that is used to make the measurement. Depending on the perception of focus, it may be overestimated (early recognition) or underestimated (late recognition) by the depth of field of the focal plane. Variations in top and bottom plane focal positions within the field of view but outside the actual sampling site require top and bottom plane focal positions to found at the sampling site.

For oil immersion lenses with a numerical aperture (NA, usually present on the objective lens collar) in the range of 1.2 to 1.4, the depth of field would be around 0.5 µm. For fluorescence microscopy, this value varies with the wavelength of the light that is observed – it becomes

smaller towards the blue end of the spectrum and may reach about 0.8 µm towards the red end of the spectrum for a NA 1.2 objective lens. Using 0.5 µm thickness for the focal plane in the middle of the visible spectrum, it will only be possible to measure section thickness to within 1 µm of its true value. Note that the error is independent of section thickness, i.e. the thicker the section is, the smaller will the relative error become. One µm of a five µm section amounts to a possible error of 20% – the same measurement error will amount to only 2% if the sections are 50 µm thick. There are no firm rules for the size of the error that would be acceptable, but a thickness of the sections close to 20 µm or more after processing is often aimed for (e.g., Wirenfeldt et al., 2003), i.e. a possible error of 5% for objective lenses with NAs around 1.3.

We have assumed that at least the top and bottom surfaces of the section are flat planes. This is rarely, if ever, the case. Depending on the embedding and sectioning, the surfaces will appear more or less uneven (Carlo and Stevens, 2011; Helander, 1983). The least unevenness will be present in sections from tissues embedded in hard media (Helander, 1983) cut with very sharp knives, while vibratome-cut fresh tissue may exhibit massive surface irregularities. When we approach the irregular section surface with the focal plane, not all parts of the field of view will appear “in focus” at the same time. Deciding at which time to start the measurement will be second possible source of error. Ideally, section thickness should be measured within the areas of the section, in which we will place the probes (Figure 42), e.g., if Disectors are used, it is best to measure somewhere within the counting frame that represents the Disector. Alas, this generates yet another problem. At the location of the probe, there may not be anything stained present at the top or bottom of the section if a stain is used that is selective for structures that occupy only little of the volume of the section. Fortunately, even unstained tissue is not invisible or impossible to focus. Background staining and/or light diffraction of different tissue components will give the tissue a fine texture when in focus (Figure 43). Also, small dust particles that have settled on the section during processing should be contained in the depth of field together with focused tissue at the top of the section (Figure 43).

Measuring thickness is not trivial, and a single measurement may consume more time than assessing a sampling location for probe-feature interactions. To limit the effort, section thickness does not need to be measured at each sampling location. In the same ways in which we can select representative samples of the tissue to assess with probes, we may

select a representative sample of thickness measurements, i.e. we can decide to measure thickness at, e.g., every 2nd, 5th or 10th sampling location. Also, as will be discussed in section 11.6, it may not make sense at all to measure thickness at locations at which we do not obtain counts with the probes. Thickness measurements at these locations may therefore be skipped. Efforts are being made automate section thickness measurements (Elozory et al., 2012).

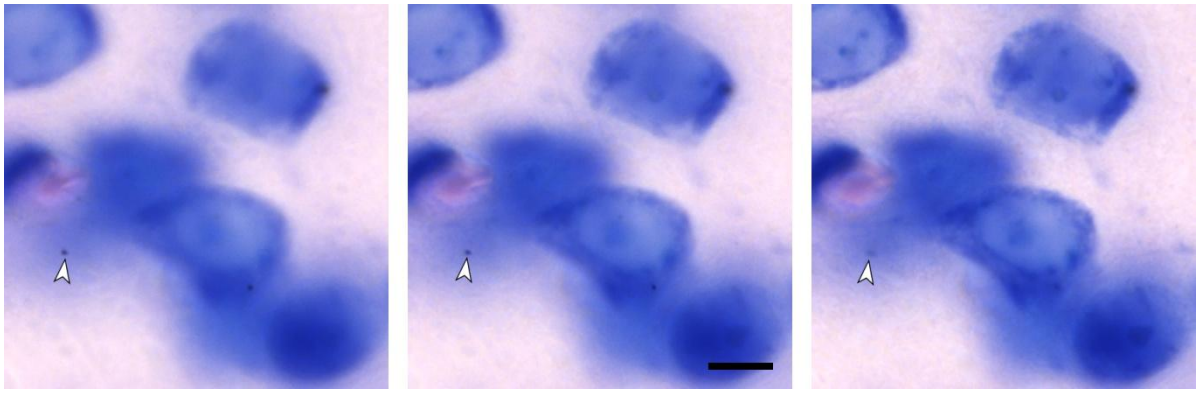

Figure 43 **Finding the top focal plane of a section.** The three images of Giemsa stained methacrylate embedded rat cortex were acquired with 0.6 µm steps along the z-axis. The left image is clearly out of focus. Some cellular images appear sharp and the background becomes finely structured in the center image – possibly representing the early recognition of the top focal plane in Figure 42. While sharpness further increases in the right image, a fine grain of dust on the surface of the section (arrowheads) is now out of focus. The depth of field should be able to contain both dust and tissue at the very top of the section. This image possibly represents late recognition of the top focal plane in Figure 42.

11.2 Using the entire thickness of the section

Guard zones have been included in the optical Disector on technical or observational rather than mathematical grounds (Andersen and Gundersen, 1999). The technical reason of the loss of objects close to the section surfaces can be investigated. Although sections prepared using different cutting techniques do exhibit different degrees of surface artifacts, these artifacts may not interfere with the probing of the tissue. In contrast to observations in vibratome sections (Andersen and Gundersen, 1999), Carlo and Stevens (2011) found that although cryostat sections do show a surface roughness, the surface can be considered smooth when the nucleus was the counting unit. Because surface irregularities were on a

much smaller scale than the size of the counting unit, they would be unlikely to result in the loss of counting units. If there is no loss, there is no technical reason to include a guard zone.

There are two observational reasons that may require guard zones. First, as already mentioned in section 8.4 *Guard zones*, the correct identification of counting units that have a complex shape may require a deep guard zone. At least with regard to the appearance of nuclei or nucleoli in the nervous system, we are lucky. Light microscopically, their correct identification is not a problem. Complexly shaped counting units, like vascular branch points are more demanding and at least their valence, i.e. whether a vessel divides into two, three or more smaller vessel, is likely to require a deep guard zone.

The second observational reason for the inclusion of guard zones is the lost caps problem that already haunted Abercrombie's method (Konigsmark, 1970). Lost caps refer to parts of the counting unit that are too small to be recognized in a section, like the topmost cap of a nucleus just included in the section at its lower surface. If this cap is not recognized, i.e. lost from the count, an underestimate will result. This will occasionally be the case if we include the lower surface of the section in the Disector probe. After the cap was lost in one section, the remainder of the counting unit will be located in the following section. Would the nucleus (or any other counting unit) be immediately recognized when we look at the top of this section or not? If we do, the unit is indeed lost. If we do not, it will be counted in the following section. As pointed out by West (2012b), the theoretical chance that the exact point is hit at which the unit is not recognized in one section but recognized in the next one is negligible. Practically, this is not the case. What is recognized in the topmost focal plane of a section may be influenced by what is seen out of focus behind it – a phenomenon also known from the physical Disector (Tang et al., 2001). There is no way of telling how much of the underestimate will remain, and endless further possibilities might be discussed skewing the estimate in one or the other direction. Empirical evidence suggests that counts that are obtained with minimal guard zones may not differ from those with larger ones (Schmitz et al., 2000). To obtain the best possible numbers, it needs to be judged if errors associated with maximizing and measuring section thickness or non-homogeneous shrinkage (section 11.6 *Shrinkage*) are likely to be smaller or larger than those associated with the omission of guard zones. This call can only be made by an assessment of the material at hand.

11.3 Brightfield, fluorescence, confocal and electron microscopy

The principles of sampling and probing do not change with the type of microscopy that is being used. However, workload and technical requirements may differ dramatically. If structures of interest can be resolved reliably using traditional brightfield microscopy, there is no reason to use more advanced techniques.

E.g., single-label immunofluorescence microscopy means working in the dark with fading, sometimes less than brilliant dyes without offering clear methodological advances over brightfield immunohistochemistry. For multiple-label studies one may have to resort to fluorescence (Amrein et al., 2015), although there are alternatives (Hsu and Soban, 1982; McMenamin, 1999; Osman et al., 2013). Not all traditional epifluorescence images are as blurry as advertisements for confocal microscopes would suggest, and some red-green-blue confocal finger paintings only vaguely resemble nervous tissue. Confocal microscopy must be resorted to if the increase in z- resolution is needed, although x- and y-resolution increase less. Stack acquisition times may be very fast, but analysis of the stacks is not. Analysis will most likely require the intermittent storage of large numbers of high-resolution stacks for offline analysis. Further practical points that need attention when confocal microscopy becomes part of quantitative morphology have been discussed by Peterson (2014) and Kubínová and Janáček (2015). The confidence and routine of a laboratory to generate the high quality material necessary for quantitative assessment may well outweigh other considerations when a method of light microscopy is chosen.

The generation of a sufficient number of sampling sites has been a critical factor in many electron microscopic applications. Often each site to be sampled has been equal to one tissue block to be processed (Geinisman et al., 1996; West et al., 2009), limiting the sampling sites assessed to numbers that would be considered (too) low in other contexts. A detailed workflow for electron microscopic study can be found in West et al. (2009). Advances in the control of the stage and image acquisition techniques in electron microscopy have made it possible to collect multiple samples from large sections (Reichmann et al., 2015). Ion beam milling (Knott et al., 2008) may facilitate sample acquisition further. Because of the large depth of field of electron microscopes, there is no equivalent in traditional electron microscopy the optical application of probes in light microscopy. Physical sections need to be compared for number estimates, and isotropic probe-feature interactions for length and

surface estimates need to be guaranteed by randomizing the orientation of the tissue using the Isector or vertical sections.

11.4 Cutting, staining and coverslipping

Both the placement of optical probes within tissue section and the measurement of section thickness ask for thick sections. This request is sometimes in conflict with our abilities to cut or embed sections and, most importantly, to adequately stain the sections.

Before these problems are addressed we will quickly look at the number of series that are to be cut. Unless only one or two series are cut, prime numbers of series should be avoided. If, e.g., five series are cut, the only options are either counting in one series, i.e., at every 5th section, or in every section by pooling all series. There are no choices in-between that would allow the pooling of series *and* maintaining uniform distances between the sections that have been pooled. If, instead, six series are cut, all series could be pooled to count in every section, every 2nd series (series 1, 3 and 5, or series 2, 4 and 6) could be pooled to look at every 2nd section, every 3rd series (series 1 and 4, series 2 and 5, or series 3 and 6) could be pooled to look at every 3rd section, or just one series could be used to count in every 6th section. If it is not clear from the outset if one of the six series is sufficient to obtain a count that is “good enough”, other series may be reserved. One may try to obtain a good estimate using series 1 and reserve series 4 just in case it turns out that more sections are needed.

Cutting – Cutting presents the least problem when fixed tissue is being processed. From our own experience, it does become difficult to cut cryostat sections much thicker than 50 µm without the sections developing cracks that may interfere with subsequent processing or with the assessment under the microscope. Up to this thickness, thicker sections usually require higher (closer to 0°C) settings of cryostat temperature. Thicker sections of excellent quality can be cut in the form of frozen sections (Figure 44), i.e. a frozen tissue block is cut with a room-temperature knife. The sections melt in the form of ugly little sausages onto the knife, but unfold beautifully without rolling once they are picked up with a brush and placed into a liquid medium (cryoprotectant, buffer etc, Figure 44C and D). Depending on the speed of the movement of the knife (thicker sections – slower movement), we have cut sections up to 200 µm thick and not reached the limit of what would be feasible. Sliding microtomes are easily modified to cut frozen sections (Figure 44A). Paraffin embedding and cutting techniques have also been adjusted to allow the preparation of section up to 100 µm thick

(Feldengut et al., 2013). If the sections do not require staining because the structures are intrinsically visible, e.g., because of pigmentation, autofluorescence or the induced expression of fluorescent proteins, only the working distance of the microscope objective limits the thickness of the sections that can be assessed.

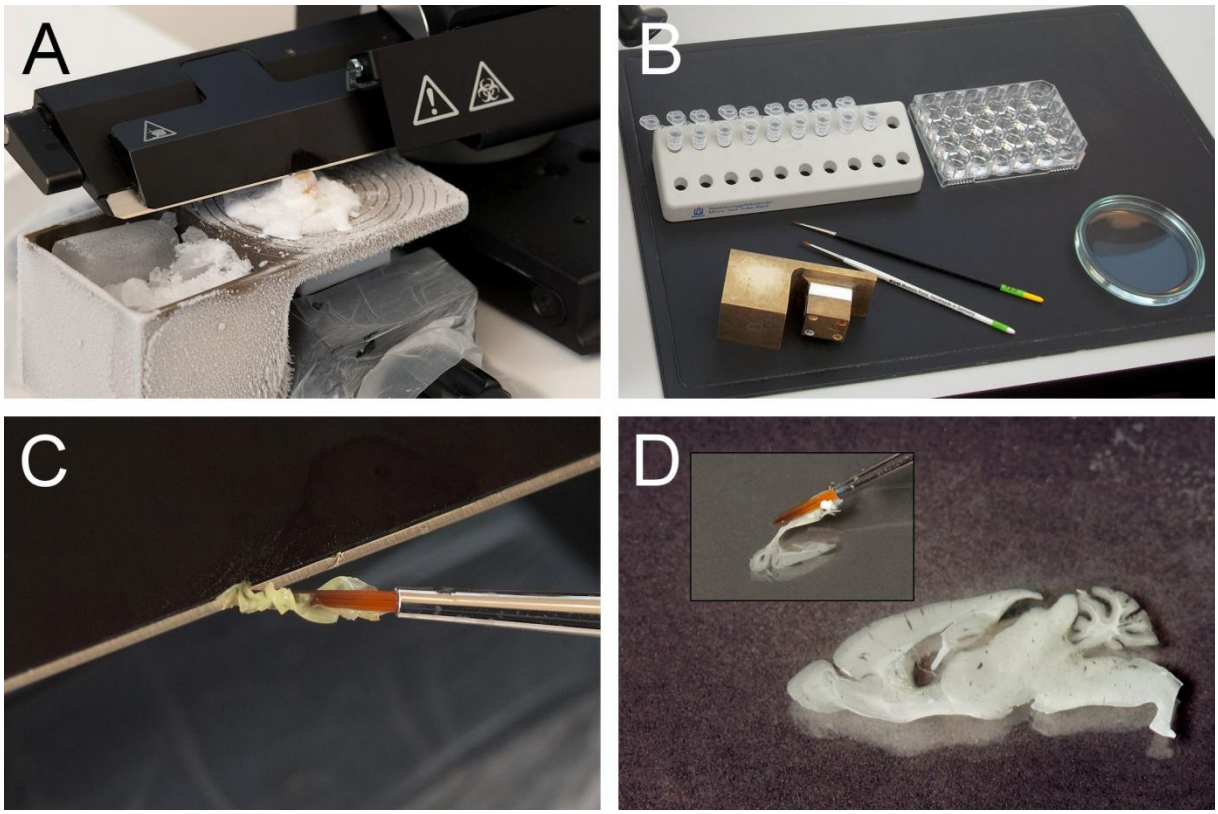

Figure 44 Preparing frozen sections on a modified sliding microtome. (**A**) A custom made freezing attachment filled with dry ice (and a few ml of alcohol to facilitate heat transfer) cools the mounted tissue. (**B**) A well plate to collect sections of a reference series in correct anatomical order and lockable tubes to collect 9 additional series for storage prior to further processing. (**C**) Initially disconcerting appearance of a freshly cut section being collected off the knife, and (**D**) unfolding of the actual section illustrated in (**C**) when placed into buffer. (Images by courtesy of Heinz Sonderegger)

Staining – Staining is not critical using many classical histological stains that are based on low molecular weight dyes that easily penetrate the tissue, such as counter-stains applied to immunohistochemically stained sections or as stand-alone Nissl stains. Using high molecular weight compounds during, e.g., immunohistochemical staining, penetration of the compounds into the tissue rapidly becomes the factor that limits the thickness of the sections that can be processed (e.g., Lyck et al., 2006; Nomura et al., 1997). While decreasing

aldehyde concentration in the fixative may increase penetration, increasing incubation times or membrane permeabilization may improve penetration less than one would expect (Torres et al., 2006). The use of zinc-based fixatives (Beckstead, 1994) instead of aldehyde-based ones has been reported to increase penetration for some surface antigens to 100 µm (Nikolajsen et al., 2016). Critical to the suitability of the material is the presence of stained structures in the center of the section. Ideally, the stained structures are evenly distributed throughout the section, which can be formally assessed by looking at the z-axis distribution of structure-probe interactions (Figure 20, see also section 11.6 *Shrinkage*).

Did every structure that we are interested in stain? Even classical Nissl counter-stains have been reported to fail staining neurons of interest that could be rendered visible immunohistochemically (Whitney et al., 2008) or to return significantly lower number estimates than those obtained from immunohistochemically stained sections (Zhu et al., 2015). In turn, markers like NeuN, which are thought to be expressed by most neurons (Mullen et al., 1992, who already defined some exceptions), may generate much lower neuron counts than classical stains (Lyck et al., 2009). The selective loss of NeuN from neurons (McPhail et al., 2004; Portiansky et al., 2006; Ünal-Cevik et al., 2004; Wu et al., 2010) is maybe not that surprising if one considers that it is a gene splicing factor (Rbfox-3; Dent et al., 2010; Kim et al., 2009; Kim et al., 2011) with phosphorylation-dependent antigenicity (Lind et al., 2005). It has been suggested that functional states of neurons may influence NeuN expression or that NeuN may even be suitable to investigate functional state (Duan et al., 2016; Maxeiner et al., 2014). The loss of marker expression rather than a loss of the cells that expressed the marker has been observed in other instances (e.g., Filice et al., 2016; Stanley and Shetty, 2004).

Using a single marker, it would be difficult to conclusively answer if all structures of interest stained. Both the loss of marker expression from otherwise persisting structures and the loss of the structures themselves may be viable and equally interesting interpretations of the data.

Coverslipping – Another way to "increase" section thickness is to avoid losing section thickness during the final steps of tissue processing. Most of the decrease in section thickness is due to the drying and dehydration of the sections prior to coverslipping, and a section cut at 40 to 60 µm in the cryostat may only be 12 to 15 µm thick once it has been processed and is coverslipped (Bonthius et al., 2004; Carlo and Stevens, 2011). An efficient

way to conserve section thickness is the coverslipping with aqueous mounting media that avoid the drying and dehydration steps (Bonthius et al., 2004), of which the drying step seems to impact section thickness most. Although commercial preparations are quite expensive in comparison to organic solvent-based media, the expense may be small in comparison to the remainder of a project. Also, it is possible to prepare rather inexpensive aqueous media like Moviol or Apathy's mounting medium. Using aqueous coverslipping media, section thicknesses close to the value set during cutting can be obtained. In addition to providing thick sections that minimize errors in the estimation of section thickness with ample room for the placement of optical probes, the wider spacing of structures along the z-axis may also make it easier to identify structure-probe interactions. E.g., densely packed cells are easier to count in a Disector if they retain their size along the z-axis instead of shrinking into a stack of pancakes.

11.5 Missing sections

The loss of sections during the cutting of the tissue is usually obvious. The number of the series concerned and the position of the missing section should be noted. The loss of sections or parts thereof during later stages of, in particular, free-floating tissue processing may be less obvious. Sections may inadvertently stick to sieves or tube caps, become entangled in brushes or be siphoned into oblivion during the pipetting of solutions. The loss of sections can and should be a rare event. It is a very rare event in many labs, and it is recommended to consult one of these for advice or training if the loss of material is considered a problem. A problem that needs to be solved when dealing with sections lost unknowingly during tissue processing is to find out if sections are actually missing. Although requiring some modest bookkeeping, the easiest way would be noting how many sections have been cut and collected in each series. The reference series already mentioned in the preceding section may also allow determining the (at least approximate) position of the lost section in the series by comparing the appearance of the sections that are present with each other and with the appearance of the sections in the reference series.

How are missing sections dealt with? If sections or parts of sections are lost in some animals of a group but not in all of them, the estimates obtained from the animals in which material is lost will turn out smaller than the estimates in the animals in which all material was available. The group mean will decrease and the group variance will increase. If the amount

of lost material in control and experimental groups is similar and if the position of the lost material is random, one may decide not to do anything and just accept increased group variances. Increased variances will decrease chances of detecting group differences, but, again, provided that losses are comparable in controls and experimental, differences in the means should be preserved. If the position of the missing section can be determined, it is valid to interpolate a value for the missing section from the adjacent sections. Some stereology software packages do contain the necessary routines. While this will restore a valid mean, some increase in variance will remain. If the purpose of the study is descriptive, i.e. the actual mean and not a difference between two means carry the important information, interpolation will be necessary. If the lost section carries a significant amount of signal or if multiple sections have been lost, it may be advisable to discard the series from further analysis. There are no firm rules when this needs to be done.

11.6 Shrinkage

Shrinkage is the nemesis of all quantitative morphological methods. The measurement of section thickness is required because of shrinkage and shrinkage may or may not influence the counts that are returned by the probes that are applied to the tissue.

Volume – Implicit in the term shrinkage is a change in the volume of the tissue that is investigated. Dorph-Petersen et al. (2005) provide a detailed decription how gross volume changes can be monitored prior to sectioning. If shrinkage is known, pre-shrinkage estimates may be calculated to facilitate the comparison of data from different laboratories. Note that different regions of the brain and different compartments of each region may shrink differently (discussed in Dorph-Petersen and Lewis, 2011).

Number – Estimates of number, which is a dimensionless parameter, are sometimes presented as the only estimates that cannot be affected by tissue shrinkage. Raisins in dough are the example that is used to illustrate why this should be so. The dough may rise or collapse/shrink, but the numbers of raisins in the dough will stay the same. Raisin density, of course, will change. At least this is true for physical Fractionator designs. They do not require measurements along any of the tissue axes, and they assess the entire thickness of the sections. Other number estimators require a measurement along the z-axis to obtain section thickness. The thickness of the fully processed section may, e.g., depend on the number of cells found at the location of the probe. This is akin to dough that collapses and that now has

bumps in the places in which there were many raisins. We see a differential shrinkage; section thickness is dependent on the local number of structures that one is interested in. This type of differential shrinkage can be accounted for. The thickness sampling fractions need to be corrected by calculating a number-weighted section thickness (Dorph-Petersen et al., 2001). If one cell is counted at a sampling site at which thickness is measured to 10 µm and five cells are counted at a site at which thickness is measured to 20 µm, number-weighted mean section thickness would be (1 × 14 µm + 5 × 20 µm) / (1 + 5) = 19 µm. A full example of the calculations has been provided by Dorph-Petersen et al. (2001). If there is no differential shrinkage, estimates based on number-weighted thickness will not differ from estimates based on the mean thickness calculated across sampling sites without reference to cells numbers. Number-weighted thickness can therefore be the default choice even if there is no suspicion of differential shrinkage.

The attentive reader may have noticed that sampling was discussed with regard to sections, the selection of sampling sites within sections, but not with regard to the placement of probes along the thickness of the section. It is again the guard zones that prevent using the entire thickness of the sections. Because parts of the section are excluded from probing, no practical approach to sample the depth of the sections was developed for the optical Disector (Andersen and Gundersen, 1999). This has raised concerns about the z-axis representativeness of the sampling sites. If, e.g., a 10 µm deep Disector probe is always placed below a two µm thick guard zone, all samples come from a depth of the section between two and 12 µm. Other depths of the tissue do not contribute to the sample, and the sample is no longer guaranteed to be representative of the tissue. If the section is affected by shrinkage in its most unpleasant form – non-homogeneous shrinkage (Dorph-Petersen et al., 2001) – it may not be representative. Non-homogeneous shrinkage refers to shrinkage that has different effects at different depth of the section. Tissue close to the surface may shrink more than tissue in the center of the section. If all samples come from one depth range an error is generated that amounts to the difference of the counts at this specific depth and the counts that would have been obtained if depth had been sampled representatively. Depending on the material, object density differences along the depth of the section may be substantial. All preparation techniques may be affected (Baryshnikova et al., 2006; Gardella et al., 2003), including plastic sections (Hatton and von Bartheld, 1999), and the effect may well differ from laboratory to laboratory (compare Figure 20 with data in

Hatton and von Bartheld (1999)). The severity of the problem is difficult to judge, and a formalized solution to this problem has not been proposed yet. The selection of guard zone depth is typically based on an assessment of the z-axis distribution of objects, which can show problems if they are present. Also, it is usually more problematic to obtain a section thickness sufficient to place a decently sized probe within the tissue than to decide where the probe should be placed along the depth of section that are substantially thicker than required. Many probes will cover much of the tissue depth thereby including much of the variation, if present, in local densities. Omitting guard zones and assessing the entire depth of the section is another option already discussed in section 11.2 *Using the entire thickness of the section*.

Area and length – Although shrinkage refers to changes of volume, it is usually used in a global sense – tissue blocks shrink during dehydration prior to embedding and cutting, and the thickness of sections shrinks when they dry after mounting. That does not mean that all components within the tissue block or section shrink to the same degree. The situation becomes even more equivocal concerning areas and length. As two- and one-dimensional measures, they do not occupy a volume. Therefore, length and area estimates *may* change, but they *do not have to* change when the tissue shrinks. For example, cell membranes may crumple instead of decreasing in size, and axons or capillaries may take a more tortuous course rather than shorten. If shrinkage is accompanied by any changes and, if so, how large they are is impossible to determine retrospectively. The bottom line is that *shrinkage should be avoided* if it is important that volume, surface or length estimates are close to the *in vivo* value of the parameter.

11.7 Estimate Presentation

Sometimes it is unintentionally funny to see how quantitative methodology is presented. Cells were counted “according to the principles of unbiased stereology” or “following StereoInvestigator” (the name of a leading stereology software package). This may mean no more than something being counted somehow in a series of sections and multiplied afterwards by the section sampling fraction. Alas, the systematic and maybe even representative sampling of sections is at least a start. But one wonders if “according to the principles of good immunocytochemistry” or “following Bond-III” (an automatic staining machine) would have passed review just as easily. The specific methods must of course be

mentioned, and credit should be given to the researchers that went through the trouble of devising them. Beyond that, everything that is needed to replicate a study is needed in the presentation of an estimate, and information useful to evaluate the quality of the outcomes would at least be helpful. Recommendations made here overlap to a large degree with those made by others (Dorph-Petersen and Lewis, 2011; Schmitz and Hof, 2005; West, 2012a)

Definitions – Even perfect sampling and measuring are no guarantee for similar estimates across different laboratories. The definition of the region of interest provides ample opportunity for estimates to diverge. This is because of the requirement to assess the entire region and not just some "representative" sections. It is in the parts of the region in which its anatomical appearance starts to diverge from the typical in which differences in definitions may create differences in the estimates. Divergent estimates may all be methodologically valid, but only for the definitions that have been used. It may appear that the demands of the methods that have been introduced here negate anatomical expertise (section 5.1 *The "representative" section*). They do, in fact, require more of it by way of the need to define the region of interest when it is not looking foolproof. Definitions will impact on the outcomes. At the same time, outcomes should be replicable. Definitions should therefore be part of the presentation of the estimates. In quantitative studies, definitions that are comprehensive have been provided for, e.g., the hippocampus in mouse (West et al., 1978) and human (West and Gundersen, 1990), the rat striatum (Oorschot, 1996), the rat and primate amygdala (Carlo et al., 2010; Chareyron et al., 2011; Mills Schumann and Amaral, 2005) or the rat and human entorhinal cortex (Mulders et al., 1997; West and Slomianka, 1998). Outside the quantitative realm, the amply illustrated literature produced during the heydays of descriptive morphology in the 1950's to 1980's may be helpful in defining regions beyond their typical appearance in today's often single, stamp-sized image. If definitions at a sufficient level of detail have been published, they can, of course, be referred to.

Sampling parameters and estimate precision – Sampling parameters are essential to replicate the outcomes of a study. For each stage at which a sample was drawn, the associated parameters should be provided: the frequency at which sections were sampled, the distance between sampling sites within sections and the size of the samples (e.g., the dimensions of the Disector samples or the radius of the Spaceballs). Collectively, the parameters represent the amount of probe that was used to obtain the quantitative signal.

Conceptually, these parameters are similar to the amount of an antibody or of an RNA-probe used to obtain an immunocytochemical or hybridization signal.

In immunocytochemical or *in situ* hybridization studies, images are used to illustrate the strength of the signal that was obtained. In quantitative studies, the strength of the signal corresponds to the number of probe-feature interactions at the sampling sites. By themselves these numbers do not provide any information on the quality of the quantitative signal, but they do allow a judgment of the workload associated with a particular sampling scheme. It is not the strength of the signal *per se* that determines its quality, but the signal-to-noise ratio. An immunohistochemically stained cell that looks almost black may be showing strong signal, but it is not very helpful if the background is almost as dark. Images would provide this information in an immunocytochemical study. The *CE* is a measure of the noise that sampling will generate in a quantitative morphological study. The critical CE^2/CV^2 can be calculated if the standard deviations, means and *CE*s are reported. It may be sufficient to comment on the overall size (>0.5 or <0.5) of this ratio to document that the estimation procedures are likely to be “good enough”. While other parameters, such as the number of sections or sampling sites, are informative, an anatomically informed reader/reviewer can at least guess at them quite well based on the sampling parameters.

Exemplary descriptions of methodology can be found in, e.g., Woodruff-Pak (2006), Carlo et al. (2010), Stranahan et al. (2012) or Filice et al. (2016). If manuscript space is at a high premium, shortcuts to the inclusion of parameters related to quantitative procedures are possible if they have been documented before and if they are referenced.

12 Perspectives in Quantitative Morphology

Design-based stereological methods are the only ones that currently provide statistically valid estimates about quantitative morphological features in tissue sections when the actual determination (e.g., counting *all* cells) is not practically feasible. Methods to probe for the basic parameters volume, surface, length and number are mature and may not improve significantly. Improvements are to be expected primarily with regard to sampling efficiency, the estimation of the precision of estimates (e.g., Hall and Ziegel, 2011; Mattfeldt, 2006; Mattfeldt, 2011; Ziegel et al., 2010) and the automation of probing (e.g., Hansen et al., 2011).

Tissue sectioning will always be an impediment to the three-dimensional visualization and understanding of brain structure. Of course, if it is not necessary to section the tissue and if it is practically feasible to determine instead of estimate a parameter, these methods will become obsolete. Not surprisingly, several methods, e.g., BABB (Dodt et al., 2007), Sca*l*e and Sca*l*eS (Hama et al., 2015; Hama et al., 2011), 3DISCO (Ertürk et al., 2012), CLARITY (Chung et al., 2013), PACT/RIMS/PARS (Yang et al., 2014), 3D BrainCV (Wu et al., 2014) or SWITCH (Murray et al., 2015), have already been developed to avoid tissue sectioning. It will be interesting to see how far these approaches can be pushed in terms of the size of tissue that can be processed, the range of the traditional probes (antibodies, RNA probes etc) that can be applied to them and the accessibility to a wide community of scientists. Only processing power would seem to set the limit for the automated analysis of image stacks of entire brain structures, and first quantitative data have been presented (Silvestri et al., 2015). If there is not sufficient power for a determination and in the course of validation, statistically representative probes will have to be efficiently drawn and analyzed, retaining the probing and sampling of the methods that have been introduced here. As far as the estimation of number is concerned, several approaches for the automated detection of objects in three dimensions have been proposed (e.g, Bjornsson et al., 2008; Chinta and Wasser, 2012; Dumitriu et al., 2012; Fish et al., 2008). Approaches that have been tested for their detection of cells did however not return robust values when compared to a human observer. Depending on the setting, true positive detection rates ranged between 33 and 99%, and false negative detections ranged between 3.6 and 82% for the best approach tested (Schmitz et al., 2014). While the generation of algorithms that produce accurate numbers under a

defined set of condition seems possible, little is gained if they time and again need to be calibrated to different stains or experimental conditions.

Anyone interested in quantitative morphology must be keenly looking forward to the problems of the automated quantitative morphological assessment of tissues being solved. Unfortunately, solutions do not seem to be just around the corner, and it is definitely not advisable to defer the generation of pilot data for the next grant application in the hope that they may arrive in time. Quantitative morphology as introduced here will still be with us for some time to come. Also, much of the brain is still quantitatively uncharted territory. A basic parameter like the number of neurons in a volume unit of the somatosensory cortex had to be estimated for a recent study that aimed at a simulation of cortical function (Markram et al., 2015). Even well-done descriptive quantitative studies rarely appear in the highest impact-factor publications, but their impact may well outlast many studies that do, because, sooner or later, qualitative and quantitative parameters will have to integrate to provide an understanding of function of neural systems.

13 Acknowledgements

I have been very lucky with kind, fair and fostering teachers that I have met with starting in Primary School and continuing until today. I remember the names of Sabine Schloo, Dr. Lothar Schwetlik, Herbert Blaffert and Jürgen Captuller with particular regard. I would like to thank you very much! I am grateful for the friendship and support of Mark J. West.

I sincerely thank for the contributions in the form of time, encouragement, comments, criticism, numbers, or images by Irmgard Amrein, Gery Barmettler, Peter Groscurth, Stanley E. Lazic, Lukas Sommer, R. Maarten van Dijk and David P. Wolfer.

14 References

Abercrombie M. 1946. Estimation of nuclear population from microtome sections. Anat Rec 94:239-247.

Abusaad I, MacKay D, Zhao J, Stanford P, Collier DA, Everall IP. 1999. Stereological estimation of the total number of neurons in the murine hippocampus using the optical disector. J Comp Neurol 408:560-566.

Acer N, Çankaya MN, İşçi Ö, Baş O, Çamurdanoğlu M, Turgut M. 2010. Estimation of cerebral surface area using vertical sectioning and magnetic resonance imaging: a stereological study. Brain Res 1310:29-36.

Ackman JB, Siddiqi F, Walikonis RS, LoTurco JJ. 2006. Fusion of microglia with pyramidal neurons after retroviral infection. J Neurosci 26:11413-11422.

Adiguzel E, Duzcan SE, Akdogan I, Tufan AC. 2003. A simple low-cost method for two dimensional microscopic measuring and stepping on the microscopic plate. Neuroanatomy 2:6-8.

Amrein I, Nosswitz M, Slomianka L, van Dijk RM, Engler S, Klaus F, Raineteau O, Azim K. 2015. Septo-temporal distribution and lineage progression of hippocampal neurogenesis in a primate (*Callithrix jacchus*) in comparison to mice. Front Neuroanat 9:85.

Amrein I, Slomianka L, Lipp H-P. 2004. Granule cell number, cell death and cell proliferation in the dentate gyrus of wild-living rodents. Eur J Neurosci 20:3342-3350.

Andersen BB, Gundersen HJG. 1999. Pronounced loss of cell nuclei and anisotropic deformation of thick sections. J Microscopy 196:69-73.

Artacho-Pérula E, Roldán-Villalobos R, Cruz-Orive LM. 1999. Application of the fractionator and vertical slices to estimate total capillary length in skeletal muscle. J Anat 195:429-437.

Azim K, Fiorelli R, Zweifel S, Hurtado-Chong A, Yoshikawa K, Slomianka L, Raineteau O. 2012. 3-Dimensional examination of the adult mouse subventricular zone reveals lineage-specific microdomains. PLoS ONE 7:e49087.

Baddeley A, Dorph-Petersen K-A, Vedel Jensen EB. 2006. A note on the stereological implications of irregular spacing of sections. J Microscopy 222:177-181.

Baddeley AJ, Gundersen HJG, Cruz-Orive LM. 1986. Estimation of surface area from vertical sections. J Microscopy 142:259-276.

Baldwin SA, Gibson T, Callihan CT, Sullivan PG, Palmer E, Scheff SW. 1997. Neuronal cell loss in the CA3 subfield of the hippocampus following cortical contusion utilizing the optical disector method for cell counting. J Neurotrauma 14:385-398.

Baquet ZC, Williams D, Brody J, Smeyne RJ. 2009. A comparison of model-based (2D) and design-based (3D) stereological methods for estimating cell number in the substantia nigra pars compacta (SNpc) of the C57BL/gJ mouse. Neuroscience 161:1082-1090.

Baryshnikova LM, von Bohlen und Halbach O, Kaplan S, von Bartheld CS. 2006. Two distinct events, section compression and loss of particles ("lost caps"), contribute to z-axis distortion and bias in optical disector counting. Microsc Res Tech 69:738-756.

Basler L. 2011. Estimatoren von Messfehlern in quantitativen morphologischen Untersuchungen. Thesis, Dr med, University of Zürich.

Beckstead JH. 1994. A simple technique for preservation of fixation-sensitive antigens in paraffin-embedded tissues. J Histochem Cytochem 42:1127-1134.

Ben Abdallah NM-B, Slomianka L, Vyssotski AL, Lipp HP. 2010. Early age-related changes in adult hippocampal neurogenesis in C57 mice. Neurobiol Aging 31:151-161.

Bjornsson CS, Lin G, Al-Kofahi Y, Narayanaswamy A, Smith KL, Shain W, Roysam B. 2008. Associative image analysis: a method for automated quantification of 3D multi-parameter images of brain tissue. J Neurosci Meth 170:165-178.

Bonilha L, Kobayashi E, Cendes F, Li LM. 2003. Effects of method and MRI slice thickness on entorhinal cortex volumetry. Neuroreport 14:1291-1295.

Bonthius DJ, McKim R, Koele L, Harb H, Karacay B, Mahoney J, Pantazis NJ. 2004. Use of frozen sections to determine neuronal number in the murine hippocampus and neocortex using the optical disector and optical fractionator. Brain Res Protocols 14:45-57.

Brændgaard H, Gundersen HJG. 1986. The impact of recent stereological advances on quantitative studies of the nervous system. J Neurosci Meth 18:39-78.

Buckmaster PS, Dudek FE. 1997. Neuron loss, granule cell axon reorganization, and functional changes in the dentate gyrus of epileptic kainate-treated rats. J Comp Neurol 385:385-404.

Buffon G-LL. 1777. Essais d'Arithmetique Morale. L'Histoire Naturelle, générale et particulière, avec la description du Cabinet du Roi, suppléments IV. Paris: L'Imprimerie Royale. p 46-109.

Carlo CN, Stefanacci L, Semendeferi K, Stevens CF. 2010. Comparative analyses of the neuron numbers and volumes of the amygdaloid complex in old and new world primates. J Comp Neurol 518:1176-1198.

Carlo CN, Stevens CF. 2011. Analysis of differential shrinkage in frozen brain sections and its implications for the use of guard zones in stereology. J Comp Neurol 519:2803-2810.

Chareyron LJ, Banta Lavenex P, Amaral DG, Lavenex P. 2011. Stereological analysis of the rat and monkey amygdala. J Comp Neurol 519:3218-3239.

Chen S, Buckmaster PS. 2005. Stereological analysis of forebrain regions in kainate-treated epileptic rats. Brain Res 1057:141-152.

Chinta R, Wasser M. 2012. Three-dimensional segmentation of nuclei and mitotic chromosomes for the study of cell divisions in live Drosophila embryos. Cytometry 81:52-64.

Chung K, Wallace J, Kim S-Y, Kalyanasundaram S, Andalman AS, Davidson TJ, Mirzabekov JJ, Zalocusky KA, Mattis J, Denisin AK and others. 2013. Structural and molecular interrogation of intact biological systems. Nature 497:332-337.

Clarke PGH. 1992. How inaccurate is the Abercrombie correction factor for cell counts? Trends Neurosci 15:211-212.

Coggeshall RE, Lekan HA. 1996. Methods for determining numbers of cells and synapses: a case for more uniform standards of review. J Comp Neurol 364:6-15.

Coimbra JP, Collin SP, Hart NS. 2014. Topographic specializations in the retinal ganglion cell layer of Australian passerines. J Comp Neurol 522:3609-3628.

Collan Y. 1998. Alternatives for morphometric and stereologic analysis in toxicopathology. Toxicol Lett 102-103:393-397.

Cotel M-C, Bayer TA, Wirths O. 2008. Age-dependent loss of dentate gyrus granule cells in APP/PS1KI mice. Brain Res 1222:207-213.

Cotel M-C, Jawhar S, Christensen DZ, Bayer TA, Wirths O. 2012. Environmental enrichment fails to rescue working memory deficits, neuron loss, and neurogenesis in APP/PS1KI mice. Neurobiol Aging 33:96-107.

Cruz-Orive LM. 1990. On the empirical variance of a fractionator estimate. J Microscopy 160:89-95.

Cruz-Orive LM. 1994. Toward a more objective biology. Neurobiol Aging 15:377-388.

Cruz-Orive LM. 1997. Stereology of single objects. J Microscopy 186:93-107.

Cruz-Orive LM. 1999. Precision of Cavalieri sections and slices with local errors. J Microscopy 193:182-198.

Cruz-Orive LM. 2004. Precision of the fractionator from Cavalieri designs. J Microscopy 213:205-211.

Cruz-Orive LM. 2006. A general variance predictor for Cavalieri slices. J Microscopy 222:158-165.

Cruz-Orive LM, Geiser M. 2004. Estimation of particle number by stereology: an update. J Aerosol Med 17:197-212.

Cruz-Orive LM, Gelšvartas J, Roberts N. 2014. Sampling theory and automated simulations for vertical sections, applied to human brain. J Microscopy 253:119-150.

Cruz-Orive LM, Weibel ER. 1981. Sampling designs for stereology. J Microscopy 122:235-257.

Curtis K, Rollins M, Carryl H, Kimberley B, van Rompay KK, Abel K, Burke MW. 2014. Reduction of pyramidal and immature hippocampal neurons in pediatric simian immunodeficiency virus infection. Neuroreport 25:973-978.

Delaloye S, Kraftsik R, Kuntzer T, Barakat-Walter I. 2009. Does the physical disector method provide an accurate estimation of sensory neuron number in rat dorsal root ganglia? J Neurosci Meth 176:290-297.

Dent MA, Segura-Anaya E, Alva-Medina J, Aranda-Anzaldo A. 2010. NeuN/Fox-3 is an intrinsic component of the neuronal nuclear matrix. FEBS Lett 584:2767-2771.

Dodt HU, Leischner U, Schierloh A, N. J, Mauch CP, Deininger K, Deussing JM, Eder M, Zieglgänsberger W, Becker K. 2007. Ultramicroscopy: three-dimensional visualization of neuronal networks in the whole mouse brain. Nature Methods 4:331-336.

Dorph-Petersen K-A, Delevich KM, Marcsisin MJ, Zhang W, Sampson AR, Gundersen HJG, Lewis DA, Sweet RA. 2009. Pyramidal neuron number in layer 3 of primary auditory cortex of subjects with schizophrenia. Brain Res 1285:42-57.

Dorph-Petersen K-A, Nyengaard JR, Gundersen HJG. 2001. Tissue shrinkage and unbiased stereological estimation of particle number and size. J Microscopy 204:232-246.

Dorph-Petersen K-A, Pierri JN, Perel JM, Sun Z, Sampson AR, Lewis DA. 2005. The influence of chronic exposure to antipsychotic medications on brain size before and after tissue fixation: a comparison of haloperidol and olanzapine in macaque monkeys. Neuropsychopharmacology 30:1649-1661.

Dorph-Petersen K-A, Pierri JN, Wu Q, Sampson AR, Lewis DA. 2007. Primary visual cortex volume and total neuron number are reduced in schizophrenia. J Comp Neurol 501:290-301.

Dorph-Petersen KA, Lewis DA. 2011. Stereological approaches to identifying neuropathology in psychosis. Biol Psychiat 69:113-126.

Duan W, Zhang YP, Hou Z, Huang C, Zhu H, Zhang C-Q, Yin Q. 2016. Novel insights into NeuN: from neuronal marker to splicing regulator. Mol Neurobiol 53:1637-1647.

Dumitriu D, Berger SI, Hamo C, Hara Y, Bailey M, Hamo A, Grossman YS, Janssen WG, Morrison JH. 2012. Vamping: stereology-based automated quantification of fluorescent puncta size and density. J Neurosci Meth 209:97-105.

Ebbeson SOE, Tang D. 1965. A method for estimating the number of cells in histological sections. J Microscopy 84:449-464.

Elozory DT, Kramer KA, Chaudhuri B, P. BO, Goldgof DB, Hall LO, Mouton PR. 2012. Automatic section thickness determination using an absolute gradient focus function. J Microscopy 248:145-259.

Ertürk A, Becker K, Jährling N, Mauch CP, Hojer CD, Egen JG, Hellal F, Bradke F, Sheng M, Dodt HU. 2012. Three-dimensional imaging of solvent-cleared organs using 3DISCO. Nature Protocols 7:1983-1995.

Evans GW. 1917. Cavalieri's theorem in his own words. Amer Math Monthly 24:447-451.

Feldengut S, Del Tredici K, Braak H. 2013. Paraffin sections of 70–100 μm: a novel technique and its benefits for studying the nervous system. J Neurosci Meth 215:241-244.

Filice F, Vörckel KJ, Sungur AÖ, Wöhr M, Schwaller B. 2016. Reduction in parvalbumin expression not loss of the parvalbumin-expressing GABA interneuron subpopulation in genetic parvalbumin and shank mouse models of autism. Mol Brain 9:10.

Fish KN, Sweet RA, Deo AJ, Lewis DA. 2008. An automated segmentation methodology for quantifying immunoreactive puncta number and fluorescence intensity in tissue sections. Brain Res 1240:62-72.

Forsman CA, Lindh B, Elfvin LG, Hallman H. 1989. Measurements of the DNA amount in mono- and binucleate cells in the celiac superior mesenteric ganglion of the guinea pig. Anat Embryol 179:587-590.

Galbraith W. 1955. The optical measurement of depth. Quart J Microc Sci 96:285-288.

García-Fiñana M, Cruz-Orive LM. 2004. Improved variance prediction for systematic sampling on R. Statistics 38:243-272.

Gardella D, Hatton WJ, Rind HB, Rosen GD, von Bartheld S. 2003. Differential tissue shrinkage and compression in the z-axis: implications for optical disector counting in vibratome-, plastic and cryosections. J Neurosci Meth 124:45-59.

Gardi JE, Nyengaard JR, Gundersen HJG. 2006. Using biased image analysis for improving unbiased stereological number estimation – a pilot simulation study of the smooth fractionator. J Microscopy 222:242-250.

Gardi JE, Nyengaard JR, Gundersen HJG. 2008. Automatic sampling for unbiased and efficient stereological estimation using the proportionator in biological studies. J Microscopy 230:108-120.

Geinisman Y, Gundersen HJG, van der Zee E, West MJ. 1996. Unbiased stereological estimation of the total number of synapses in a brain region. J Neurocytol 25:805-819.

Glagolev AA. 1933. On geometrical methods of quantitative mineralogical analysis of rocks [in Russian]. Trans Inst Econ Min Moscow 59:1-47.

Glagolev AA. 1955. On geometrical methods of quantitative mineralogical analysis of rocks (translated from Russian by J.B. Sykes). Harwell, Berkshire: Atomic Energy Research Establishment. 51 p.

Glaser EM, Wilson PD. 1998. The coefficient of error of optical fractionator population size estimates: a computer simulation comparing three estimators. J Microscopy 192:163-171.

Gokhale AM. 1990. Unbiased estimation of curve length in 3-D using vertical slices. J Microscopy 159:133-141.

Gokhale AM, Evans RA, Mackes JL, Mouton PR. 2004. Design-based estimation of surface area in thick tissue sections of arbitrary orientation using virtual cycloids. J Microscopy 216:25-31.

Goldschmidt RB, Steward O. 1992. Retrograde regulation of neuronal size in the entorhinal cortex: consequences of the destruction of dentate granule cells with colchicine. Restorative Neurol Neurosci 3:335-343.

Gritti I, Henny P, Galloni F, Mainville L, Mariotti M, Jones BE. 2006. Stereological estimates of the basal forebrain cell population in the rat, including neurons containing choline

acetyltransferase, glutamic acid decarboxylase or phosphate-activated glutaminase and colocalizing vesicular glutamate transporter. Neuroscience 143:1051-1064.

Guillery RW, Herrup K. 1997. Quantification without pontification: choosing a method for counting objects in sectioned tissues. J Comp Neurol 386:2-7.

Gundersen HJG. 1977. Notes on the estimation of the numerical density of arbitrary profiles: the edge effect. J Microscopy 147:219-223.

Gundersen HJG. 1986. Stereology of arbitrary particles. J Microscopy 143:3-45.

Gundersen HJG. 1988. The nucleator. J Microscopy 151:3-21.

Gundersen HJG. 2002. The smooth fractionator. J Microscopy 207:191-210.

Gundersen HJG, Bagger P, Bendtsen TF, Evans SM, Korbo L, Marcussen N, Møller A, Nielsen K, Nyengaard JR, Pakkenberg B and others. 1988a. The new stereological tools: disector, fractionator, nucleator and point sampled intercepts and their use in pathological research and diagnosis. APMIS 96:857-881.

Gundersen HJG, Bendtsen TF, Korbo L, Marcussen N, Møller A, Nielsen K, Nyengaard JR, Pakkenberg B, Sørensen FB, Vesterby A and others. 1988b. Some new, simple and efficient stereological methods and their use in pathological research and diagnosis. APMIS 96:379-394.

Gundersen HJG, Jensen EB. 1987. The efficiency of systematic sampling in stereology and its prediction. J Microscopy 147:229-263.

Gundersen HJG, Jensen EBV, Kieu K, Nielsen J. 1999. The efficiency of systematic sampling in stereology – reconsidered. J Microscopy 193:199-211.

Gundersen HJG, Østerby R. 1981. Optimizing sampling efficiency of stereological studies in biology: or "Do more less well!". J Microscopy 121:65-73.

Hall P, Ziegel J. 2011. Distribution estimators and confidence intervals for stereological volumes. Biometrika 98:417-431.

Hama H, Hioki H, Namiki K, Hoshida T, Kurokawa H, Ishidate F, Kaneko T, Akagi T, Saito T, Saido T and others. 2015. ScaleS: an optical clearing palette for biological imaging. Nature Neurosci 18:1518-1529.

Hama H, Kurokawa H, Kawano H, Ando R, Shimogori T, Noda H, Fukami K, Sakaue-Sawano A, Miyawaki A. 2011. Scale: a chemical approach for fluorescence imaging and reconstruction of transparent mouse brain. Nature Neurosci 14:1481-1488.

Hansen LV, Nyengaard JR, Andersen JB, Jensen EBV. 2011. The semi-automatic nucleator. J Microscopy 242:206-215.

Hatton WJ, von Bartheld CS. 1999. Analysis of cell death in the trochlear nucleus of the chick embryo: calibration of the optical disector counting method reveals systematic bias. J Comp Neurol 409:169-186.

Hedreen JC. 1998a. Lost caps in histological counting methods. Anat Rec 250:366-372.

Hedreen JC. 1998b. What was wrong with the Abercrombie and empirical cell counting methods? A review. Anat Rec 250:373-380.

Helander KG. 1983. Thickness variations within individual paraffin and glycol methacrylate sections. J Microscopy 132:223-227.

Hosseini-Sharifabad M, Nyengaard JR. 2007. Design-based estimation of neuronal number and individual neuronal volume in the rat hippocampus. J Neurosci Meth 162:206-214.

Howard CV, Reed MG. 2010. Unbiased Stereology. Three-dimensional Measurements in Microscopy. Liverpool: QTP Publications. 277 p.

Hsu SM, Soban E. 1982. Color modification of diaminobenzidine (DAB) precipitation by metallic ions and its application for double immunohistochemistry. J Histochem Cytochem 30:1079-1082.

Hykšová M, Kalousová A, Saxl I. 2012. Early history of geometric probability and stereology. Image Anal Stereol 31:1-16.

Jørgensen A-MB, Marner L, Pakkenberg B. 2008. No change in total length of white matter fibers in Alzheimer's disease. Neuroscience 157:878-883.

Kaplan S, Canan S, Aslan H, Ünal B, Sahín B. 2001. A simple technique to measure the movements of the microscope stage along the x and y axes for stereological methods. J Microscopy 203:321-325.

Karadaglić D, Wilson T. 2008. Image formation in structured illumination wide-field fluorescence microscopy. Micron 39:808-818.

Kiêu K. 1997. Three lectures on systematic geometric sampling. Aarhus: Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus. 99 p.

Kim KK, Adelstein RS, Kawamoto S. 2009. Identification of Neuronal Nuclei (NeuN) as Fox-3, a new member of the Fox-1 gene family of splicing factors. J Biol Chem 284:31052-31061.

Kim KK, Kim YC, Adelstein RS, Kawamoto S. 2011. Fox-3 and PSF interact to activate neural cell-specific alternative splicing. Nucleic Acids Res 39:3064-3078.

Kleder M. 2005. WGS84 ellipsoidal earth plotted at 4 pixels per degree.

Knott G, Marchman H, Wall D, Lich D. 2008. Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. J Neurosci 28:2959-2964.

Kolinko Y, Cendelin J, Kralickova M, Tonar Z. 2016. Smaller absolute quantities but greater relative densities of microvessels are associated with cerebellar degeneration in Lurcher mice. Front Neuroanat 10:35.

Konigsmark BW. 1970. Methods for the counting of neurons. In: Nauta WJH, Ebbeson SOE, editors. Contemporary research methods in neuroanatomy. Berlin: Springer. p 315-340.

Korbo L, Pakkenberg B, Ladefoged O, Gundersen HJG, Arlien-Søborg P, Pakkenberg H. 1990. An efficient method for estimating the total number of neurons in rat brain cortex. J Neurosci Meth 31:93-100.

Kordower JH. 2000. Making the counts count: the stereology revolution. J Chem Neuroanat 20:1-2.

Korkmaz M, Tümkaya L. 1997. Estimation of the section thickness and optical disector height with a simple calibration method. J Microscopy 187:104-109.

Kruskal W, Mosteller F. 1979. Representative sampling, II: scientific literature, excluding statistics. Int Stat Rev 47:111-127.

Kubínová L, Janácek J. 1998. Estimating surface area by the isotropic fakir method from thick slices cut in an arbitrary direction. J Microscopy 191:201-211.

Kubínová L, Janáček J. 2015. Confocal stereology: an efficient tool for measurement of microscopic structures. Cell Tissue Res 360:13-28.

Larsen JO, Gundersen HJG, Nielsen J. 1998. Global spatial sampling with isotropic virtual planes: estimators of length density and total length in thick, arbitrarily oriented sections. J Microscopy 191:238-248.

Lind D, Franken S, Kappler J, Jankowski J, Schilling K. 2005. Characterization of the neuronal marker NeuN as a multiply phosphorylated antigen with discrete subcellular localization. J Neurosci 79:295-302.

Løkkegaard A, Nyengaard JR, West MJ. 2001. Stereological estimates of number and length of capillaries in subdivisions of the human hippocampal region. Hippocampus 11:726-740.

Lyck L, Dalmau Santamaria I, Pakkenberg B, Chemnitz J, Schrøder HD, Finsen B, Gundersen HJG. 2009. An empirical analysis of the precision of estimating the numbers of neurons and glia in human neocortex using a fractionator-design with sub-sampling. J Neurosci Meth 182:143-156.

Lyck L, Ljelsing J, Søndergaard Jensen P, Lambertsen KL, Pakkenberg B, Finsen B. 2006. Immunohistochemical visualization of neurons and specific glial cells for stereological application in the porcine neocortex. J Neurosci Meth 152:229-242.

Markram H, Muller E, Ramaswamy S, Reimann MW, Abdellah M, Sanchez CA, Ailamaki A, Alonso-Nanclares L, Antille N, Arsever S and others. 2015. Reconstruction and simulation of neocortical microcircuitry. Cell 163:456-492.

Matérn B. 1985. Estimating area by dot counts. In: Lanke J, Lindgren G, editors. Contributions to Probability and Statistics in Honour of Gunnar Blom. Lund: Department of Mathematical statistics, Lund University. p 243-257.

Matheron G. 1965. Les variables régionalisées et leur estimation. Paris: Masson et Cie. 305 p.

Matheron G. 1971. The Theory of Regionalized Variables and Its Applications. Paris: École National Supérieure des Mines de Paris. 211 p.

Mattfeldt T. 2006. Prediction of the variance of stereological volume estimates from systematic sections using computer-intensive methods. J Microscopy 222:166-176.

Mattfeldt T. 2011. A brief introduction to computer-intensive methods, with a view towards applications in spatial statistics and stereology. J Microscopy 242:1-9.

Maxeiner S, Glassmann A, Kao HT, Schilling K. 2014. The molecular basis of the specificity and cross-reactivity of the NeuN epitope of the neuron-specific splicing regulator, Rbfox3. Histochem Cell Biol 141:43-55.

Mayhew TM. 1996. How to count synapses unbiasedly and efficiently at the ultrastructural level: proposal for a standard sampling and counting protocol. J Neurocytol 25:793-804.

McMenamin PG. 1999. Distribution and phenotype of dendritic cells and resident tissue macrophages in the dura mater, leptomeninges, and choroid plexus of the rat brain as demonstrated in wholemount preparations. J Comp Neurol 405:553-562.

McPhail LT, McBride CB, McGraw J, Steeves JD, Tetzlaff W. 2004. Axotomy abolishes NeuN expression in facial but not rubrospinal neurons. Exp Neurol 185:182-190.

Melvin N, Poda D, Sutherland RJ. 2007. A simple and efficient alternative to implementing systematic random sampling in stereological designs without a motorized microscope stage. J Microscopy 228:103-106.

Mertz J. 2011. Optical sectioning microscopy with planar or structured illumination. Nature Methods 8:811-819.

Miki T, Harris SJ, Wilce PA, Takeuchi Y, Bedi KS. 2003. Effects of alcohol exposure during early life on neuron numbers in the rat hippocampus. I. Hilus neurons and granule cells. Hippocampus 13:388-398.

Miki T, Harris SJ, Wilce PA, Takeuchi Y, Bedi KS. 2004. Effects of age and alcohol exposure during early life on pyramidal cell numbers in the CA1-CA3 region of the rat hippocampus. Hippocampus 14:124-134.

Mills Schumann C, Amaral DG. 2005. Stereological estimation of the number of neurons in the human amygdaloid complex. J Comp Neurol 491:320-329.

Morgan JT, Barger N, Amaral DG, Schumann CM. 2014. Stereological study of amygdala glial populations in adolescents and adults with autism spectrum disorder. PLoS ONE 9:e110356.

Mouton PR, Gokhale AM, Ward NL, West MJ. 2002. Stereological length estimation using spherical probes. J Microscopy 206:54-64.

Mulders WHAM, West MJ, Slomianka L. 1997. The numbers of neurons in the presubiculum, parasubiculum and entorhinal area of the rat. J Comp Neurol 385:83-94.

Mullen RJ, Buck CR, Smith AM. 1992. NeuN, a neuronal specific nuclear protein in vertebrates. Development 116:201-211.

Murray E, Cho JH, Goodwin D, Ku T, Swaney J, Kim SY, Choi H, Park Y, Park JY, Hubbert A and others. 2015. Simple, scalable proteomic imaging for high-dimensional profiling of intact systems. Cell 163:1500-1514.

Neil M, Juskaitis R, Wilson T. 1997. Method of obtaining optical sectioning by using structured light in a conventional microscope. Opt Lett 22:1905-1907.

Nikolajsen GN, Jensen MS, West MJ. 2016. A zinc fixative for 3D visualization of cerebral capillaries and pericytes. J Neurosci Meth 257:1-6.

Nissl F. 1892. Ueber die Veränderungen der Ganglienzellen am Facialiskern des Kanninchens nach Ausreissung der Nerven. Allg Z Psychiat 48:197-198.

Nomura T, Fukuda T, Aika Y, Heizmann CW, Emson PC, Kobayashi T, Kosaka T. 1997. Distribution of nonprincipal neurons in the rat hippocampus, with special reference to their dorsoventral difference. Brain Res 751:64-80.

Nyengaard JR, Gundersen HJG. 1992. The isector: a simple and direct method for generating isotropic, uniform random sections from small specimens. J Microscopy 165:427-431.

Oorschot DE. 1996. Total number of neurons in the neostriatal, pallidal, subthalamic, and substantia nigral nuclei of the rat basal ganglia: a stereological study using the cavalieri and optical disector methods. J Comp Neurol 366:580-599.

Osman TA, Oijordsbakken G, Costea DE, Johannessen AC. 2013. Successful triple immuno-enzymatic method employing primary antibodies from same species and same immunoglobulin subclass. Eur J Histochem 57:e22.

Pakkenberg B, Gundersen HJG. 1997. Neocortical neuron number in humans: effect of sex and age. J Comp Neurol 384:312-320.

Peterson DA. 2014. High-resolution estimation of multiple cell populations in tissue using confocal stereology. In: Cornea A, Conn PM, editors. Fluorescence Microscopy. Super-Resolution and Other Novel Techniques. London: Academic Press. p 171-184.

Phinney AL, Calhoun ME, Woods AG, Deller T, Jucker M. 2004. Stereological analysis of the reorganization of the dentate gyrus following entorhinal cortex lesions in mice. Eur J Neurosci 19:1731-1740.

Portiansky EL, Barbeito CG, Gimeno EJ, Zuccolilli GO, Goya RG. 2006. Loss of NeuN immunoreactivity in rat spinal cord neurons during aging. Exp Neurol 202:519-521.

Porzionato A, Macchi V, Parenti A, de Caro R. 2009. Morphometric analysis of infant and adult medullary nuclei through optical disector method. Anat Rec 292:1619-1629.

Pover CM, Coggeshall RE. 1991. Verification of the disector method for counting neurons, with comments on the empirical method. Anat Rec 231:573-578.

Reichmann F, Painsipp E, Holzer P, Kummer D, Bock E, Leitinger G. 2015. A novel unbiased counting method for the quantification of synapses in the mouse brain. J Neurosci Meth 240:13-21.

Rezze CJ. 1966. Células binucleadas no núcleo do nervo hipoglosso humano através dos grupos etários. Arq Neuro-Psiquiatr 24:247-256.

Ribeiro AACM. 2006. Size and number of binucleate and mononucleate superior cervical ganglion neurons in young capybaras. Anat Embryol 211:607-617.

Roberts N, Cruz-Orive LM, Reid NMK, Brodie DA, Bourne M, Edwards RHT. 1993. Unbiased estimation of human body composition by the Cavalieri method using magnetic resonance imaging. J Microscopy 171:239-253.

Rudow G, O'Brien R, Savonenko AV, Resnick SM, Zonderman AB, Pletnikova O, Marsh L, Dawson TM, Crain BJ, West MJ and others. 2008. Morphometry of the human substantia nigra in ageing and Parkinson's disease. Acta Neuropathol 115:461-470.

Saltykov SA. 1946. The method of intersections in metallography (in Russian). Zavodskaja laboratorija 12:816-825.

Saltykov SA. 1974. Stereometrische Metallographie. Leipzig: VEB Deutscher Verlag für Grundstoffindustrie. 397 p.

Saper CB. 1996. Any way you cut it: a new journal policy for the use of unbiased counting methods. J Comp Neurol 364:5.

Schmitz C. 1998. Variation of fractionator estimates and its prediction. Anat Embryol 198:371-397.

Schmitz C, Born M, Dolezel P, Rutten BPF, de Saint-Georges L, Hof PR, Korr H. 2005. Prenatal protracted irradiation at very low dose rate induces severe neuronal loss in rat hippocampus and cerebellum. Neuroscience 130:935-948.

Schmitz C, Dafotakis M, Heinsen H, Mugrauer K, Niesel A, Popken GJ, Stephan M, van de Berg WDJ, von Hörsten S, Korr H. 2000. Use of cryostat sections from snap-frozen nervous tissue for combining stereological estimates with histological, cellular, or molecular analyses on adjacent sections. J Chem Neuroanat 20:21-29.

Schmitz C, Eastwood BS, Tappan SJ, Glaser JR, Peterson DA, Hof PR. 2014. Current automated 3D cell detection methods are not a suitable replacement for manual stereologic cell counting. Front Neuroanat 8:27.

Schmitz C, Hof PR. 2000. Recommendations for straightforward and rigorous methods of counting neurons based on a computer simulation approach. J Chem Neuroanat 20:93-114.

Schmitz C, Hof PR. 2005. Design-based stereology in neuroscience. Neuroscience 130:813-831.

Silvestri L, Paciscopi M, Soda P, Biamonte F, Iannello G, Frasconi P, Pavone FS. 2015. Quantitative neuroanatomy of all Purkinje cells with light sheet microscopy and high-throughput image analysis. Front Neuroanat 9:68.

Slomianka L, West MJ. 1987. Asymmetry in the hippocampal region specific for one of two closely related species of wild mice. Brain Res 436:69-75.

Slomianka L, West MJ. 2005. Estimators of the precision of stereological estimates: an example based on the CA1 pyramidal cell layer of rats. Neuroscience 136:757-767.

Smith CS, Guttman L. 1953. Measurement of internal boundaries in three-dimensional structures by random sectioning. Trans AIME (J Metals) 197:81-87.

Stanley DP, Shetty AK. 2004. Aging in the rat hippocampus is associated with widespread reductions in the number of glutamate decarboxylase-67 positive interneurons but not interneuron degeneration. J Neurochem 89:204-216.

Stark AK, Toft MH, Pakkenberg H, Fabricius K, Eriksen N, Pelvig DP, Møller M, Pakkenberg B. 2007. The effect of age and gender on the volume and size distribution of neocortical neurons. Neuroscience 150:121-130.

Sterio DC. 1984. The unbiased estimation of number and sizes of arbitrary particles using the disector. J Microscopy 134:127-136.

Stranahan AM, Jiam NT, Stocker AM, Gallagher M. 2012. Aging reduces total neuron number in the dorsal component of the rodent prefrontal cortex. J Comp Neurol 520:1318-1326.

Syková E, Mazel T, Hasenöhrl RU, Harvey AR, Simonová Z, Mulders WHAM, Huston JP. 2002. Learning deficits in aged rats related to decrease in extracellular volume and loss of diffusion anisotrpy in hippocampus. Hippocampus 12:269-279.

Tang Y, Nyengaard JR. 1997. A stereological method for estimating to total length and size of myelin fibers in human brain white matter. J Neurosci Meth 73:193-200.

Tang Y, Nyengaard JR, de Groot DMG, Gundersen HJG. 2001. Total regional and global number of synapses in the human brain neocortex. Synapse 41:258-273.

Thompson WR. 1932. The geometric properties of microscopic configurations. I. general aspects of projectometry. Biometrika 24:21-26.

Torres EM, Meldrum A, Kirik D, Dunnett SB. 2006. An investigation of the problem of two-layered immunohistochemical staining in paraformaldehyde fixed sections. J Neurosci Meth 158:64-74.

Ünal-Cevik I, Kilinc M, Gürsoy-Özdemir Y, Gurer G, Dalkara T. 2004. Loss of NeuN immunoreactivity after cerebral ischemia does not indicate neuronal cell loss: a cautionary note. Brain Res 1015:169-174.

Waite Boyce R, Dorph-Petersen K-A, Lyck L, Gundersen HJG. 2010. Design-based stereology: introduction to basic concepts and practical approaches for estimation of cell number. Toxicol Pathol 38:1011-1025.

West MJ. 2012a. Basic Stereology for Biologists and Neuroscientists. Cold Spring Harbor: Cold Spring Harbor Laboratory Press. 203 p.

West MJ. 2012b. Estimating object number in biological structures. Cold Spring Harb Protoc 2012:1049-1066.

West MJ. 2012c. The precision of estimates in stereological analyses. Cold Spring Harb Protoc 2012:937-949.

West MJ, Bach G, Søderman A, Ledet Jensen J. 2009. Synaptic contact number and size in stratum radiatum CA1 of APP/PS1E9 transgenic mice. Neurobiol Aging 30:1756-1776.

West MJ, Coleman PD. 1996. How to count. Neurobiol Aging 17:503.

West MJ, Coleman PD, Flood DG. 1988. Estimating the number of granule cells in the dentate gyrus with the disector. Brain Res 448:167-172.

West MJ, Danscher G, Gydesen H. 1978. A determination of the volumes of the layers of the rat hippocampal region. Cell Tissue Res 188:345-359.

West MJ, Gundersen HJG. 1990. Unbiased stereological estimation of the number of neurons in the human hippocampus. J Comp Neurol 296:1-22.

West MJ, Slomianka L. 1998. Total number of neurons in the layers of the human entorhinal cortex. Hippocampus 8:69-82.

West MJ, Slomianka L, Gundersen HJG. 1991. Unbiased stereological estimation of the total number of neurons in the subdivisions of rat hippocampus using the optical fractionator. Anat Rec 231:482-497.

Whitney ER, Kemper TL, Rosene DL, Bauman ML, Blatt GJ. 2008. Calbindin-D28k is a more reliable marker of human Purkinje cells than standard Nissl stains: a stereological experiment. J Neurosci Meth 168:42-47.

Wirenfeldt M, Dalmau I, Finsen B. 2003. Estimation of absolute microglial cell numbers in mouse fascia dentata using unbiased and efficient stereological cell counting principles. Glia 44:129-139.

Woodruff-Pak DS. 2006. Stereological estimation of Purkinje neuron number in C57BL/6 mice and its relation to associative learning. Neuroscience 141:233-243.

Wu J, He Y, Yang Z, Guo C, Luo Q, Zhou W, Chen S, Li A, Xiong B, Jiang T and others. 2014. 3D BrainCV: simultaneous visualization and analysis of cells and capillaries in a whole mouse brain with one-micron voxel resolution. Neuroimage 87:199-208.

Wu K-L, Li Y-Q, Tabassum A, Lu W-Y, Aubert I, Wong CS. 2010. Loss of neuronal protein expression in mouse hippocampus after irradiation. J Neuropathol Exp Neurol 69:272-280.

Xavier-Vidal R. 2010. Disector Z-axis mechanical method for stereology. An Acad Bras Cienc 82:539-544.

Yamamura T, Barker JM, Balthazart J, Ball GF. 2011. Androgens and estrogens synergistically regulate the expression of doublecortin and enhance neuronal recruitment in the song system of adult female canaries. J Neurosci 31:9649-9657.

Yang B, Treweek JB, Kulkarni RP, Deverman BE, Chen CK, Lubeck E, Shah S, Cai L, Gradinaru V. 2014. Single-cell phenotyping within transparent intact tissue through whole-body clearing. Cell 158:945-958.

Zhang W, Thamattoor AK, LeRoy C, Buckmaster PS. 2015. Surviving mossy cells enlarge and receive more excitatory synaptic input in a mouse model of temporal lobe epilepsy. Hippocampus 25:594-604.

Zhang W, Yamawaki R, Wen X, Uhl J, Diaz J, Prince DA, Buckmaster PS. 2009. Surviving hilar somatostatin interneurons enlarge, sprout axons, and form new synapses with granule cells in a mouse model of temporal lobe epilepsy. J Neurosci 29:14247-14256.

Zhu Y, Liu F, Zou X, Torbey M. 2015. Comparison of unbiased estimation of neuronal number in the rat hippocampus with different staining methods. J Neurosci Meth 254:73-79.

Ziegel J, Baddeley A, Dorph-Petersen K-A, Vedel Jensen EB. 2010. Systematic sampling with errors in sample locations. Biometrika 97:1-13.

15 Curriculum Vitae

Name: Lutz Slomianka

Birth: 27/05/1961, Flensburg, Germany

Citizenship: Danish

Address: Anatomisches Institut, Universität Zürich, Winterthurerstr. 190, 8057 Zürich, Switzerland

e-mail: slomianka@anatom.uzh.ch

Education

1987 Graduated, cand. scient., Neurobiology, Aarhus University, Denmark

1980 Graduated, majors: biology, chemistry, Altes Gymnasium, Flensburg, Germany

Employment

2006- Lecturer, ETH Zürich, Switzerland

2001- Research Associate/Lecturer/e-learning Specialist, Department of Anatomy and Studiendekanat Vorklinik (Medizin), University of Zürich, Switzerland

1994 - 2001 Lecturer, tenured since 1997, Department of Anatomy and Human Biology, University of Western Australia, Australia

1994 Research Associate / Lecturer, Institute of Anatomy, University of Rostock, Germany

1992 - 1994 Postdoctoral Fellow, Laboratory of Neuroanatomy, Department of Neurosciences, Johns Hopkins University School of Medicine, Baltimore, USA

1987 - 1992 Research and Teaching Associate, Institute of Anatomy, Aarhus University, Denmark

Teaching Experience

Awards and Commendations

2010, 2015 Medical Teacher of the Year Vorklinik, Fachverein Medizin der Universität Zürich

1997, 2000 Commendation for Excellence in Teaching, University of Western Australia Student Guild

University of Zürich (UZH) – Eidgenössische Technische Hochschule Zürich (ETH), Switzerland

2013 - Human Biology II, lectures, laboratory classes (Faculty of Medicine, UZH)

2012 - Anatomy and Physiology, lectures, laboratory classes (ETH)

2005 - 2012 Anatomy and Physiology, lectures, laboratory classes (joint UZH/ETH)

2002 - 2012 Molecular Cell Biology, Human Biology I & II, lectures, laboratory classes, web support (Faculty of Medicine, UZH)

Faculty of Medicine, University of Western Australia, Perth, Australia

2000 - 2001 Foundations of Normal Systems - lectures, tutorials and laboratories, course coordinator of anatomy component in 2001

1995 - 2001 Coordinated Neurosciences - lectures, course coordinator of anatomy component in 2001

1995 - 2000 Macroscopic Human Anatomy - lectures, dissections and microscopy classes, course coordinator of the histology component

1995 - 1999 Medical Biomorphology - lectures, dissections, microscopy classes, course coordinator

Faculty of Science, University of Western Australia, Perth, Australia

1995 - 2000 Advanced Neuroscience, lectures, practical classes, journal club

1995 - 1999 Human Ultrastructural Organization, lectures and practical classes

1997 - 1998 General Histology, lectures and microscopy classes

1995 Human Organs and Systemslectures and microscopy classes

1995 Anatomy and Human Biology, seminars

average annual time-tabled contact hours (1995-2001): 240

University of Rostock, Germany

1994 General and Systems Histology

Johns Hopkins University School of Medicine, Baltimore, USA

1993 Neuroscience - dissection

Aarhus University, Denmark

1991-1992 Tissue Histology

1992 Neurodissection

1986 Chordate Zoology

Administrative

Department of Anatomy and Human Biology, University of Western Australia, Perth, Australia

1997, 1999 Departmental Executive Committee

1996 – 1998, 2000 Honours Degree Committee (chair in 2000)

1996 - 1998 Postgraduate Committee

Funding

2011 University of Zürich Infrastructure Grant, Stereology and structured Illumination, CHF 160.000

1999-2001 Australian Research Council SPIRIT Grant (Strategic Partnership with Industry Research and Training), The effect on amyloid deposition of the human apolipoprotein E isoforms in transgenic mice, AU$ 304,000 (co-investigator)

1998-1999 Medical Research Fund of Western Australia, The effect on amyloid deposition of the human apolipoprotein E isoforms in transgenic mice - a model for Alzheimer's disease, AU$ 69,000 (co-investigator)

1996 National Health and Medical Research Council Equipment Grant, AU$ 46,000 (co-investigator)

1996-1998 National Health and Medical Research Council, Hippocampal pathways using vesicular zinc and receptors mediating zinc effects, AU$ 129,000 (principal investigator)

1995-1996 Theodore and Vada Stanley Foundation, Morphometric and neural connectivity characteristics of fetal brains at genetic risk for bipolar disorder, US$ 100,000 (principal investigator)

1992 Aarhus University Research Foundation, Development of the Zinc-Containing System in Hippocampus and Visual Cortex, DKR 30,000

1989-1992 Candidate Stipend, Scientific Training Program in Biomedical Sciences, Aarhus University, DKR 550,000

1989 NATO Science Fellowship, Neurons of Origin of Zinc-Containing Boutons, Danish Research Council, DKR 100,000

1987 Aarhus University Research Foundation, Developmental Neurotoxicity of Toluene, DKR 15,000

1986 Research Scholarship, Quantitative Comparative Neurobiology, Aarhus University Research Foundation, DKR 60,000

Publications – Peer-Reviewed Research

Final parentheses: Web of Science citations up to March 2016, current H-Index: 21, times cited: 3162

49. van Dijk RM, Huang SH, **Slomianka L**, Amrein I (2016) Taxonomic separation of hippocampal networks: principal cell populations and adult neurogenesis. Front Neuroanat, 10:22 DOI: 10.3389/fnana.2016.00022 (0)

48. van Dijk RM, Lazic SE, **Slomianka L**, Wolfer DP, Amrein I (2015) Large-scale phenotyping links adult hippocampal neurogenesis to the reaction to novelty. Hippocampus, in press DOI: 10.1002/hipo.22548 (0)

47. Huang S, **Slomianka L**, Farmer AJ, Kharlamova AV, Gulevich RG, Herbeck YE, Trut LN, Wolfer DP, Amrein I (2015) Selection for tameness, a key behavioral trait of domestication, increases adult hippocampal neurogenesis in foxes. Hippocampus, in press DOI: 10.1002/hipo.22420 (1)

46. Amrein I, Nosswitz M, **Slomianka L**, van Dijk RM, Engler S, Klaus F, Raineteau O, Azim K (2015) epto-temporal distribution and lineage progression of hippocampal neurogenesis in a primate (Callithrix jacchus) in comparison to mice. Front Neuroanat, 9:85 doi: 10.3389/fnana.2015.00085 (0)

45. Hick M, Herrmann U, Weyer SW, Mallm J-P, Tschäpe J-A, Borgers M, Mercken M, Roth FC, Draguhn A, **Slomianka L**, Wolfer DP, Korte M, Müller UC (2015) Acute function of secreted amyloid precursor protein fragment APPsα in synaptic plasticity. Acta Neuropathol 129:21-37 DOI: 10.1007/s00401-014-1368-x (8)

44. Amrein I, Becker AS, Engler S Huang, S-H, Müller J, **Slomianka L**, Oosthuizen MK (2014) Adult neurogenesis and its anatomical context in the hippocampus of three mole-rat species. Front Neuroanat 8:39, DOI: 10.3389/fnana.2014.00039 (1)

43. **Slomianka L**, Drenth T, Cavegn N, Menges D, Lazic SE, Phalanndwa MV, Chimimba CT, Amrein I (2013) The hippocampus of the eastern rock sengi: cytoarchitecture, markers of neuronal function, principal cell numbers and adult neurogenesis. Front Neuroanat 7:34, DOI: 10.3389/fnana.2013.00034 (7)

42. Cavegn N, van Dijk RM, Menges D, Brettschneider H, Phalanndwa M, Chimimba CT, Isler K, Lipp HP, **Slomianka L**, Amrein I (2013) Habitat-specific shaping of proliferation and neuronal differentiation in adult hippocampal neurogenesis of wild rodents. Front Neurosci 7:59, DOI: 10.3389/fnins.2013.00059 (9)

41. Azim K, Fiorelli R, Zweifel S, Hurtado-Chong A, Yoshikawa K, **Slomianka L**, Raineteau O (2012) 3-dimensional examination of the adult mouse subventricular zone reveals specialized lineage-specific microdomains. PLoS ONE 7:e49087, DOI: 10.1371/journal.pone.0049087 (7)

40. Klaus F, Hauser T, Lindholm AK, Cameron HA, **Slomianka L**, Lipp H-P, Amrein I (2012) Different regulation of adult hippocampal neurogenesis in Western house mice (Mus musculus domesticus) and C57BL/6 mice. Behav Brain Res 227:340-347, DOI: 10.1016/j.bbr.2011.07.026 (9)

39. **Slomianka L**, Amrein I, Knuesel I, Sørensen JC, Wolfer DP (2011) Hippocampal pyramidal cells: the reemergence of cortical lamination. Brain Struct Funct 216:301-317, DOI: 10.1007/s00429-011-0322-0 (26)

38. Gatome CW, **Slomianka L**, Lipp H-P, Amrein I (2010) Number estimates of neuronal phenotypes in layer ll of the medial entorhinal cortex of rat and mouse. Neuroscience 170:156-165, DOI: 10.1016/j.neuroscience.2010.06.048 (14)

37. Gatome C, **Slomianka L**, Mwangi DK, Lipp H-P, Amrein I (2010) The entorhinal cortex of the Megachiroptera: A comparative study of Wahlberg's epauletted fruit bat and the Straw-coloured fruit bat, Brain Struct Funct 214:375-393 DOI: 10.1007/s00429-010-0239-z (6)

36. Amrein I, **Slomianka L** (2010) A morphologically distinct granule cell type in the dentate gyrus of the red fox correlates with adult hippocampal neurogenesis, Brain Res 1328:12-24 DOI: 10.1016/j.brainres.2010.02.075 (22)

35. Ben Abdallah NM-B, **Slomianka L**, Vyssotski AL, Lipp H-P (2010) Early age-related changes in adult hippocampal neurogenesis in C57 mice, Neurobiol Aging 31:151-161 DOI: 10.1016/j.neurobiolaging.2008.03.002 (93)

34. Klaus F, Hauser T, **Slomianka L**, Lipp H-P, Amrein I (2009) A reward increases running-wheel performance without changing cell proliferation, neuronal differentiation or cell death in the dentate gyrus of C57BL/6 mice. Behav Brain Res 204:175-181 DOI: 10.1016/j.bbr.2009.06.002 (13)

33. Dumrese C, **Slomianka L**, Ziegler U, Choi SS, Kaliad A, Fulurija A, Lu W, Berg DE, Benghezal M, Marshall B, Mittl PRE (2009) The secreted Helicobacter cysteine-rich protein A causes adherence of human monocytes and differentiation into a macrophage-like phenotype, FEBS Lett 583:1637-1643 DOI: 10.1016/j.febslet.2009.04.027 (20)

32. Ismail T, Mauerhofer E, **Slomianka L** (2008) The hippocampal region of rats and mice after a single i.p. dose of clioquinol: loss of synaptic zinc, cell death and c-Fos induction. Neuroscience 157: 697-707, DOI: 10.1016/j.neuroscience.2008.09.011 (14)

31. Betschart C, Scheiner D, Maake C, Vich M, **Slomianka L**, Fink D, Perucchini D (2008) Histomorphological analysis of the urogenital diaphragm in elderly women: a cadaver study. Int J Urogynecol 19:1477-1481, DOI: 10.1007/s00192-008-0669-9 (4)

30. Ben Abdallah NM-B, **Slomianka L**, Lipp H-P (2007) A reversible effect of X-irradiation on proliferation, neurogenesis and cell death in the dentate gyrus of adult mice. Hippocampus 17:1230-1240 DOI: 10.1002/hipo.20358 (29)

29. **Slomianka L**, West MJ (2005) Estimators of the precision of stereological estimates: An example based on the CA1 pyramidal cell layer of rats , Neuroscience 136:757-767 DOI: 10.1016/j.neuroscience.2005.06.086 (109)

28. Amrein I, **Slomianka L**, Poletaeva II, Bologova, NV, Lipp H-P (2004) Marked age- and species-dependent differences in cell proliferation and neurogenesis in the dentate gyrus of wild-living rodents, Hippocampus 14:1000-1010 DOI: 10.1002/hipo.20018 (58)

27. Amrein I, **Slomianka L**, Lipp H-P (2004) Granule cell number, cell death and cell proliferation in the dentate gyrus of wild-living rodents. Eur J Neurosci 20:3342-3350 DOI: 10.1111/j.1460-9568.2004.03795.x (50)

26. Isgor C, **Slomianka L**, Watson SJ (2004) Hippocampal mossy fiber terminal field size is differentially affected in a rat model of risk-taking behavior, Behav Brain Res 153:7-14 (11)

25. West MJ, **Slomianka L** (2001) 2-D versus 3-D cell counting - a debate. What is an optical disector? Trends Neurosci 24:374 [commentary] (na)

24. West MJ, **Slomianka L** (1998) Total number of neurons in the layers of the human entorhinal cortex. Hippocampus 8:69-82 (37; erratum 9)

23. Mulders WHAM, West MJ, **Slomianka L** (1997) Neuron numbers in the presubiculum, parasubiculum and entorhinal cortices of the rat. J Comp Neurol 385:83-94 (37)

22. **Slomianka L**, Geneser FA (1997) Postnatal development of zinc-containing cells and neuropil in the hippocampal region of the mouse. Hippocampus 7:321-340 (32)

21. **Slomianka L**, Ernst E, Østergaard K (1997) Zinc-containing neurons are distinct from GABAergic neurons in the forebrain of the rat. Anat Embryol 195:165-174, DOI: 10.1007/s004290050035 (13)

20. Donovan J, **Slomianka L** (1996) Distribution of mossy fibres in the hippocampus of two closely related species of mice. Brain Res 732:253-256, DOI: 10.1016/0006-8993(96)00695-6 (2)

19. Sørensen JC, **Slomianka L**, Christensen J, Zimmer J (1995) Zinc-containing telencephalic connections to the rat striatum: a combined Fluoro-Gold tracing and histochemical study. Exp Brain Res 105:370-382 (23)

18. Garrett B, Østerballe R, **Slomianka L**, Geneser FA(1994) Cytoarchitecture and distribution of acetylcholinesterase and zinc in the visual cortex of the Parma wallaby (Macropus parma). Brain Behav Evol 43:162-172 (9)

17. Geneser FA, Holm IE, **Slomianka L** (1993) Application of the Timm and selenium methods to the central nervous system. Protocols in Neuroscience, 50-15-1-14 (na)

16. Sørensen JC, Tønder N, **Slomianka L** (1993) Zinc-positive afferents to the septum originate from distinct subpopulations of zinc-containing neurons in the hippocampal areas: a combined Fluoro-Gold tracing and histochemical study. Anat Embryol 188:107-116 (19)

15. **Slomianka L** , Geneser FA (1993) Distribution of acetylcholinesterase in the hippocampal region of the mouse. III. The area dentata. J Comp Neurol 331:225-235 (9)

14. Garrett B, **Slomianka L** (1992) Postnatal development of zinc-containing cells and neuropil in the visual cortex of the mouse. Anat Embryol 186:487-496 (24)

13. **Slomianka L**, Rungby J, Edelfors S, Ravn-Jonsen A (1992) Late postnatal growth in the dentate area of the rat compensates for volumetric changes caused by early postnatal toluene exposure. Toxicology 74:203-208 (13)

12. **Slomianka L** (1992) Neurons of origin of zinc-containing pathways and the distribution of zinc-containing boutons in the hippocampal region of the rat. Neuroscience 48:325-352 DOI: 10.1016/0306-4522(92)90494-M (129)

11. Garrett B, Sørensen JC, **Slomianka L** (1992) Fluoro-Gold tracing of zinc-containing afferent connections in the mouse visual cortices. Anat Embryol 185:451-459 (32)

10. West MJ, **Slomianka L**, Gundersen HJG (1991) Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator. Anat Rec 231:482-497 DOI: 10.1002/ar.1092310411 (1978)

9. **Slomianka L**, Geneser FA (1991) Distribution of acetylcholinesterase in the hippocampal region of the mouse. II. Subiculum and hippocampus. J Comp Neurol 312:525-536 (16)

8. Garrett B, Geneser FA, **Slomianka L** (1991) Distribution of acetylcholinesterase and zinc in the visual cortex of the mouse. Anat. Embryol. 184:461-468 (18)

7. **Slomianka L**, Geneser FA (1991) Distribution of acetylcholinesterase in the hippocampal region of the mouse. I. Entorhinal area, parasubiculum, retrosplenial area and presubiculum. J Comp Neurol 303:339-356 (19)

6. **Slomianka L**, Danscher G, Frederickson CJ (1990) Labeling of the neurons of origin of zinc-containing pathways by intraperitoneal injections of sodium selenite. Neuroscience 38:843-854, DOI: 10.1016/0306-4522(90)90076-G (100)

5. **Slomianka L**, Edelfors S, Ravn-Jonsen A, Rungby J, Danscher G, West MJ (1990) The effect of low-level toluene exposure on the developing hippocampal region of the rat: Histological evidence and volumetric findings. Toxicology 62:189-202, DOI: 10.1016/0300-483X(90)90109-T (32)

4. **Slomianka L**, Rungby J, West MJ, Danscher G, Holst Andersen A (1989) Dose-dependent bimodal effect of low-level lead exposure on the developing hippocampal region of the rat: A volumetric study. Neurotoxicology 10:177-190 (28)

3. **Slomianka L**, West MJ (1989) A comparative quantitative study of the hippocampal region of two closely related species of wild mice: inter- and intraspecific variation in the volumes of hippocampal components. J Comp Neurol 280:544-552, DOI: 10.1002/cne.902800405 (20)

2. **Slomianka L**, West MJ (1987) Asymmetry in the hippocampal region specific for one of two closely related species of wild mice. Brain Res 436:69-75, DOI: 10.1016/0006-8993(87)91557-5 (21)

1. Rungby J, **Slomianka L**, Danscher G, Holst Andersen A, West MJ (1987) A quantitative evaluation of the neurotoxic effect of silver on the volumes of the developing rat hippocampus. Toxicology 43:261-268, DOI: 10.1016/0300-483X(87)90085-0 (26)

Publications - Book Chapters & Other Media

4. Lipp H-P, Amrein I, Slomianka L, Wolfer D (2007) Natural genetic variation of hippocampal structures and behavior - an update. in: Neurobehavioral Genetics 2nd ed, Jones BC, Mormede P (eds) Boca Raton: CRC Press, 435-448

3. Groscurth P, Lüthi N, Slomianka L (2005) Histologie Atlas auf CD-ROM, 2nd revised edition [in German, Interactive Histology Atlas]. München, Germany: Elsevier, CD ISBN 3-437-41378-3

2. Groscurth P, Lüthi N, Slomianka L (2004) Histologie Atlas auf CD-ROM [in German, Interactive Histology Atlas]. München, Germany: Elsevier, CD ISBN 3-437-41377-5

1. McMahon G, Dook J, Slomianka L (2001) Body Tissues in Focus. Perth, Australia: Duit Multimedia, CD, ISBN 1-74052-000-9

Invited Seminars / Lectures / Courses

Invited Seminars / Lectures prior to 2000 are not listed.

2013-2016 Lecturer, Undergraduate Course 'Anatomy Skills', Université de Fribourg, Switzerland

2012-2015 Lecturer, Postgraduate Course 'Understanding the brain through the hippocampus', Aarhus University, Denmark

2013 Lecturer, Course 'Basic Stereology', Karolinska Institut, Stockholm, Sweden

2012 Seminar, Morphological Phenotyping, Department of Medicine, Anatomy Unit, Université de Fribourg, Switzerland

2004-2011 Lecturer at the annual International Stereology Workshop, Woods Hole Marine Biology Laboratory, USA

2009 Lecturer/Organizer, IBRO Introductory Stereology Workshop, Capetown, South Africa

2007 Lecturer, Introductory Stereology Workshop, San Diego, USA

2005 Lecturer/Organizer, Introductory Stereology Workshop, TATA Institute of Fundamental Research, Mumbai, India

2003 Mental Health Research Institute, Seminars, University of Michigan, Ann Arbor, USA

2003 Lecturer at the International Stereology Workshop, Prague

2002 Lecturer at the International Stereology Workshop, Paris

Reviewing

Funding agencies

Natural Sciences and Engineering Research Council of Canada

National Health and Medical Research Council, Australia

New Zealand Neurological Foundation

Journals

Brain Research

Brain, Genes and Immunity

Cells, Tissues, Organs (former Acta Anatomica)

Cerebral Cortex

Experimental Neurology

Genes, Brain and Behavior

Hippocampus

Journal of Chemical Neuroanatomy

Journal of Comparative Neurology

Journal of Histochemistry and Cytochemistry

Journal of Neuroscience Methods

Neurobiology of Aging

Neuropsychopharmacology

Neuroscience

Neuroscience Letters